To be or not to be a scientist?

Arnaud Chevalier[*]
Royal Holloway and IZA

Abstract:
Employers regularly complain of a shortage of qualified scientists and advocate that to remain competitive more scientists need to be trained. However, using a survey of graduates from British universities, I report that three years after graduation less than 50% of graduates from science subjects are working in a scientific occupation.
Accounting for selection into major and occupation type, I estimate the wages of graduates and report that the wage premium of science graduates only occurs when these graduates are matched to a scientific occupation—and not because science skills are in demand in all occupations. I also provide additional evidence to assess whether science graduates are pushed or pulled into non-scientific occupations. Altogether, the evidence does not support the claim that science graduates are pulled by better conditions, financial or otherwise, into non-scientific jobs.

Keywords: science, graduate, labour market
JEL Codes: I21, J24, J44

Acknowledgement:
This paper was drafted during a placement at the Department for Innovation, Universities and Skills. The views represented in this manuscripts are the author's and do not represent the view of the DIUS. Financial support from the ESRC for their placement scheme is also gratefully acknowledged. I also want to thank Stijn Broecke, Tarja Viitanen and participants at the DWPEG meeting (Sheffield) and IZA-CEDEFOP meeting for their helpful comments. I am also indebted to the editors and two anonymous referees for their comments that substantially improved this manuscript.

[*] Chevalier is also associated with Geary Institute-Dublin; SFI-Copenhagen, ROA-Maastricht and the Lifelong Center-Queensland . Contact: arnaud.chevalier@rhul.ac.uk

I Introduction

Around the world, the business community regularly bemoans the lack of (skilful) science graduates and warns that this hampers its competitive advantage and more generally future economic growth. The claims of science skill-shortages, first made by employers' groups, have largely permeated to stakeholders, whom over the years have commissioned several reports on the (science) skill-shortage (see among others the Sainsbury review (2007) in the UK, but also the President's Council of Advisors on Science and Technology (2012) for the US and in Europe, the European Parliament (2015) for examples). The claims of permanent skill-shortages do not sit well with economic models. Cappelli (2014), for example, largely dismissed them as baseless since in a competitive labour market, any skill shortages would result either in wage increase and a subsequent increase in the supply of such skills, or in the substitution of labour for capital to reduce the demand for scarce science skills. Similar points have recently been made in reviews produced by the Council of Canadian Academies (2015) and the UK Commission for Employment and Skills (2013).

In this paper, I focus on the labour market decisions of a cohort of UK graduates observed three years after graduation. A puzzling finding is that about 50% of science graduates do not work in an occupation related to science. This large "leakage"[1] of science graduates in the labour market is often mentioned in the context of the "pull factor" exercised by other industries, especially the financial sector, offering a justification for the skill shortage in scientific occupations. This paper attempts to document the reasons for the leakage of science graduates away from scientific occupations. In particular, I investigate the wages of graduates by STEM status and

[1] Throughout the paper I will use leakage in the very specific case of science graduates not working in a scientific occupation; these graduates may nonetheless be using their scientific skills, especially those who teach science, and have large public returns.

occupation type, accounting simultaneously for the selections into field of study and into occupation, to assess the claim that non-scientific occupations poach science graduates with higher wages. To do so, I rely on an extension of the standard Heckman (1979) selection model to include two selection equations, one for field of study and one for occupation type. I complement the analysis with information on the reasons for accepting the current job to check whether science graduates are pulled or pushed into a non-scientific occupation.

A disequilibrium between demand and supply of scientists should result in a wage premium for scientists. Indeed, such a premium has been found in, among others, Chevalier (2011) and Walker and Zhu (2011) for the UK. However, simple OLS estimates of the returns to subjects are biased since individuals' unobserved characteristics are likely positively correlated with subject choice and earnings. To account for this bias, new studies such as Hastings, Neilson and Zimerman (2013) and Kirkeboen, Leuven and Mogstad (2016) rely on administrative data from Chile and Norway, respectively. Using discontinuity in the allocation of applicants into field of study, they report large returns to science majors. Indeed, Altonji, Arcidiacono and Maurel (2015) conclude their extensive review of the literature by stating "The evidence suggests that much of the effect of major on earnings is causal, with STEM and business related majors leading the way."

This paper contributes to this literature by accounting for the selection into science majors and into scientific occupations. In addition it also assesses whether any premium for studying science is due to scientific skills in general or is specific to being matched to a scientific occupation. Similarly, Kinsler and Pavan (2015) highlight the importance of the occupational match on the returns to degree using a structural model of choice of field of study and occupation. Using the U.S. Baccalaureate and Beyond Longitudinal Study, they estimate returns to science of up to 20% but only for science graduates working in an occupation related to their studies.

Rather than building up a structural model, I rely on reduced form estimations that account for selection into STEM study fields and into scientific occupations. I use data from the Longitudinal Destination of Leavers of Higher Education (LDLHE) covering a sample of UK graduates from the 2003 cohort observed in November 2006. Importantly, this survey is linked to (part of) the university application form. As such, I have detailed information on the family background of the student, which I use to identify the double selection model. Specifically, I rely on information on tuition fee status and parental occupation to identify subject choice and occupation, respectively. In both the simple OLS and selection model, I find an average wage premium of 6% to 10% for graduating with a science major. However, this premium is specific to working in scientific occupations, which contradicts claims that scientists are poached by higher wages in other occupations. Instead, I predict that science graduates working in non-scientific occupations would have experienced higher wages (3.7 to 10%) had they worked in a scientific occupation. This strongly rejects the pull hypothesis that science graduates are attracted to non-scientific occupations by higher wages. Contrary to a Roy model (1951), whereby individuals choose the field of study and occupation that maximises their expected earnings, I find that up to 40% of graduates would have higher expected earnings had they chosen a different major. But this may have to do with the inability to model taste for field of study.

The second contribution is to examine non-financial reasons that may have pushed or pulled STEM graduates towards non-scientific occupations. To this end, I exploit subjective data on the reasons to have chosen the current job, as well as satisfaction with career and regrets about subject choice. These results are consistent with science graduates being pushed into non-scientific jobs. For example, science graduates working in a non-scientific occupation are 7 percentage points less likely than STEM graduates working in scientific occupations to report being satisfied with their careers.

Thirdly, I investigate heterogeneity in these results along two dimensions: between science subject and between gender. I find a large amount of variation in earnings and leakage between subjects, with more applied subjects converting a greater fraction of graduates into scientific occupations. I find little differences in terms of returns and leakage to non-scientific occupations between graduates of Math, IT, Physics and Engineering (MIPE) and other STEM subjects, but the formers are in general less positive about their current job. Along the gender dimension, women have lower (but not significantly so) returns to studying science, especially MIPE, but also report the same reasons for having chosen their current job.

Overall, the results on financial and non-financial factors are largely inconsistent with a pull hypothesis. Moreover, the lack of returns to scientific skills outside scientific occupations combined with the large proportion of science graduates not working in scientific occupations questions the emphasis on educating more science graduates.

II Literature review

Despite the disparities of data used, the literature generally agrees on a large heterogeneity in the returns to higher education by subject, with often a large wage premium for science degrees, see Altonji et al. (2015) for a recent review. However, these estimates often do not account for the self-selection into subject and are thus likely to be biased upwards. This self-selection is partly based on observable characteristics, such as academic ability or parental background, but also on unobservable ones. Berger (1988), using the NLSY, was one of the first studies to account for selection on unobservable. The exclusion variables to determine subject choices were questionable, but Berger (1998) reported that the "self-selection bias is not overwhelming" (p424). To account for students receiving additional information about their ability and taste for the subject while at university[2], Arcidiacono (2004) uses the National Longitudinal Study of the Class of 1972 to estimate a dynamic model of major choice. He reports that the largest returns to college are found in science (20% over not going to college) but that future monetary returns do not drive the major choice.

Another identification strategy to estimate returns to majors has relied on discontinuity in admission due to subject specific cut-off at institutions. Hastings et al. (2013) and Kirkeboen et al. (2016), using university application data from Chile and Norway, respectively, and administrative records on life-time income (up to 30 years), estimate major specific returns by comparing the earnings of individuals who were just accepted to their preferred field of study with individuals who just missed out. Hastings et al. (2013) report returns to health and science degrees of up to 25% and 12%, respectively. Kirkeboen et al. (2016) reports estimates of the returns compared to the alternative subjects, finding that while medical studies graduates have high returns, science and engineering have an average pay-off around $22,000, which is half of the payoff for Law or Business degrees.

A related identification is found in Ketel, Leuven, Oosterbeek and van der Klaauw (2016). They rely on a lottery among all qualified applicants in the Netherlands to determine admission to medical schools. Admission to medical schools permanently shifts the earnings distribution to the right, with successful

[2] In the US, 50% of students originally registering in STEM do not graduate from a STEM subject (Altonji et al. (2015)). This selection is less of a concern in the UK, where switching subject at university is rare.

applicants to medical schools earning 20% to 50% more than candidates who lost the lottery. They conclude that these returns are mostly driven by monopoly rent rather than increased human capital; as such, they are specific to working in the health sector.

Kingsley and Pavan (2015) also note that the specificity of returns to a degree to working in some occupations is not unique to medical studies. Specifically, they assume that agents are endowed with two skills (math and verbal) but are uncertain about this endowment until they reach the labour market. Moreover, skills are accumulated at different rates in different majors. Using a structural model, they estimate that for science graduates the returns to math skills are specific to working in a scientific job, and that there is no general return to scientific skills. As such, science majors are riskier investments than other fields of study.

III Data and Institutional Set-up

Prospective students in the UK apply to higher education institutions in the autumn preceding the start of the academic year. The application system is centralised, and for this cohort, applicants could apply to 6 institution/subject. Based on their predicted grades at the final secondary schooling national test (A-levels or equivalent), and in some cases interviews, applicants are rejected or accepted, conditionally on reaching a given grade at A-levels. Applicants receiving more than one offer must specify the one they want to accept, plus an insurance choice in case they do not eventually meet the grade criteria to enter their first choice course. When A-levels results become available, applicants meeting the conditions are accepted to their first or insurance course. At this stage, applicants who do not satisfy their admission conditions, or who did not get any offer in the first round, can apply for courses that still have vacancies. Those positions are filled on a first come first served basis, conditional on some (usually weaker) entry conditions; some 10% of students usually gain access via clearing. In 2006 (the first year for which detailed data is available) the ratio of acceptance to applications was 18.3, with no significant difference by STEM status.

Compared to previously used data, the LDLHE has some advantages: it is larger than other graduate datasets and can be linked to administrative data. It has thus precise information on academic achievement and family background. The LDLHE was conducted in November 2006 among a random sample of higher education leavers, who typically graduated in the summer of 2003[3]. The sampled population consists of leavers from higher education who responded to an initial questionnaire, the Destination of Leavers of Higher Education (DLHE) administered by the Higher Education Statistic Agency (HESA) six months after graduation. The response rate in the Destination of Leavers of Higher Education (DLHE) reaches 75%. A sample of 55,900 of these original respondents was contacted three years after graduation by HESA to take part in the LDLHE. 24,823 responded to either a postal, phone or online questionnaire; Tipping and Taylor (2007) provide evidence in favour of the representativeness of the survey when reweighted. Item non-response on the earnings question leaves us with 19,979 observations. We then select first degree holders only, aged 18 to 25 on graduation, non-special entry students and those who are currently observed in employment[4]. This leads to a sample of 9,296 observations (See Table A1 for details on the sample selection).

Observing graduates three years after graduation limits the scope for investigation investment to post-graduate studies. This could create a selection bias if for example science graduates are the most likely to engage in post-graduate

[3] The survey only includes individuals who were UK domiciled prior to attaining higher education.
[4] Like almost all of the literature, I condition on positive earnings. Hamermesh and Donald (2008) account for selection into employment, which compresses the earning differentials by 10 to 20%.

education. Evidence from the labour force survey (2001-2015) on the subsample of graduates aged 25 to 35 suggests this is not the case; the proportions with a post-graduate qualification are not significantly different between STEM and non-STEM graduates.

I first define science graduates or STEM as all graduates from Medicine, Subject allied to Medicine, Biological Science (including Psychology and Sport Science), Veterinary/Agriculture related subject, Physical science, Mathematical and Computer science, Engineering/Technologies, and Architecture. Individuals with mixed subjects in science are also classified as science undergraduates. As such, 49% of the sample has studied a STEM subject. Alternatively, I split this group into more mathematically oriented subjects only: Mathematics/IT, Physics/Chemistry, and Engineering (MIPE), which represents just over 40% of STEM graduates, and other STEM. In the UK, very few students switch major during their studies, so graduates typically gained entry to university for the subject that they end-up graduating from. As such, it is not necessary to model any decision regarding switching major choice. Table 1, reports the distribution of subject and the gender composition within subject. While the sample is 43% male, male graduates are over-represented in STEM, especially in MIPE where they represent 72% of graduates. Within STEM subjects there are large variations in gender composition, the two extremes being subject allied to medicine, where 17% of graduates are males, and Engineering and Technology, where 85% of graduates are males.

To investigate the labour market of graduates, we define scientific occupations (using the 5-digit SOC2000 codes). This definition suffers from some arbitrariness (see note under Table 1); however, it delivers sensible results: only 5% of non-scientific graduates work in a scientific occupation. Like in Roberts (2002), who uses an alternative definition based on industry, just under half of the scientific graduates work in a scientific occupation[5]. I also report statistics for the financial and teaching sectors, since those are popular alternative careers for STEM graduates[6].

Table A2 reports a matrix of occupation mobility for science graduates at 6 months and 3 years after graduation. Of interest here is that occupations are set early on. For example, 84% of STEM graduates working in a scientific occupation 6 months after graduation are still in this occupational group 30 months later. The fraction remaining in the same occupational group between the first and second interviews are 73%, 66% and 53% for teaching, other and financial occupations, respectively. Another way of looking at mobility is that only 11% of STEM graduates in another occupation three years after graduation worked in a scientific occupation 6 months after graduation. The rest of the paper focuses on occupational choice 36 months after graduation, as more detailed information is available at this stage.

Table 1 reports the fraction of graduates, by major, in the four occupation groups defined above: scientific, finance, teaching and other. In general more vocational science graduates (health, engineering, IT, architecture) have a higher probability of remaining in a scientific occupation than graduates from more theoretical scientific subjects (Biology, Physics and Math). Subjects with a lower mathematical content, like sport sciences and psychology, have the lowest proportion of graduates in scientific occupations. Financial occupations are an alternative only for graduates from math and combined science; for other majors less than 5% of graduates work in finance. Moreover, science graduates are less likely than non-science graduates to work in the financial sector, making this sector an unlikely culprit for the labour

[5] Using Labour Force Survey data, I also identify scientific occupations (at 2 digit level) as those in which the fraction of workers with a science degree is greater than the national average (42%). These more aggregated occupations overlap with the more precisely defined one that I keep for the analysis of the DLHE. Pooling Labour Force Survey data from 2001 to 2015, keeping degree holders aged between 25 and 35, I find that only 54% of science degree holders work in a scientific occupation; consistent with the presented findings.

[6] While science graduates are likely to use their science knowledge in the education sector, we define teaching as a non-scientific occupation since most teachers are not science teachers, and the data at hand does not allow us to separate between STEM and non-STEM teachers.

shortages in scientific occupations. Teaching is a popular occupation for graduates, attracting 17% of non-science graduates and 11% of STEM graduates. This is particularly a popular occupation for sport sciences graduates (31%) but also for graduates in math, physics, biology and psychology (14-20%)[7].

LDLHE respondents self-report their annual gross pay. I recode 36 observations with an unusually high salary compared to their occupation's average earnings due to coding errors (additional zero) and drop 149 individuals who claim to earn less than the national minimum wage (assuming they worked 52 weeks a year)[8]. The distribution of annual earnings for science and other graduates in October 2006 is reported in Figure 1, where science graduates are split between MIPE and other STEM. For all three groups, the distributions have a very similar bell-shape with a long right-hand tail, but are shifted to the right for MIPE graduates and even further to the right for other STEM graduates. Kolmogorov-Smirnov tests reject that the distributions for each of the groups are identical. On average, non-science graduates earn an annual income of £21,600, while MIPE and other STEM graduates earn £23,500 and 23,800, respectively.

In Figure 2, I report the earnings of graduates by science/occupation pair to provide the first evidence whether the returns to science skills are generic or specific to working in a scientific occupation. The earnings distribution of non-STEM largely overlaps with the one of STEM not working in scientific occupation, even if a Kolmogorov-Smirnov test rejects the equality of the distributions. The earnings distribution of STEM graduates in a scientific occupation is shifted to the right and

[7] Similar conclusions are reached when using Labour Force Survey (2001-2015), and keeping graduates aged 25 to 35 only.

[8] The LDLHE does not contain detailed information on hours of work, only an indicator for part-time or full-time employment. Relying on the labour Force Survey (2001-15), I find no significant differences in hours of work between STEM and non-STEM graduates or between graduates working in scientific and non-scientific occupations. So the differences in earnings that I observe in the LDLHE are unlikely to be driven by differences in hours of work between graduates from different subjects or in different occupations.

has a much fatter right tail—this is partially driven by graduates from medical schools. The mean earnings are £21,400, £21,900 and £26,700 for non-STEM, STEM not in scientific occupations and STEM in scientific occupations, respectively. The conclusion that the earnings distributions differ by occupation type rather than by subject type are similar when splitting science graduates between other-STEM and MIPE.

Table 2 dwells further on the issue of wage differential by degree and occupation, reporting the average annual earnings by detailed subject and occupation categories. MIPE and other-STEM graduates earn more than non-science graduates, but the difference is only significant in scientific occupations where the gap is around 20%. The absence of a premium for science graduates in non-scientific occupations suggests that there is little demand for scientific skills outside scientific occupations. Since non-science occupations pay, on average, less (or not significantly more in the case of finance) than scientific occupations, this first evidence does not square well with claims that the other occupations pull STEM graduates away from scientific occupations.

The description by detailed fields of study reveals the large heterogeneity in earnings within the science group. Graduates from medical schools are the clear outliers, with average earnings of £39,000. The next best paid subjects have earnings around the £25,000 mark, and include subjects allied to medicine, mathematics, engineering and architecture. At the other end of the pay distribution, psychology, biology and sport science graduates earn less than the average non-scientist graduate. Graduates from Math, Engineering, IT and subjects allied to Medicine are the only ones who earn significantly more when working in science occupations than in other (non-financial) occupations. While the rest of the analysis groups science graduates together, as it is not possible to identify the selection into each subject, it is important to remember the heterogeneity in earnings between science majors.

IV Econometric considerations

The descriptive evidence has highlighted that earnings differ by fields of study. However, since the characteristics of students, and the attended institutions, also differ, this is not conclusive evidence that there are returns to scientific skills. In line with the literature, I first rely on Ordinary Least Square estimates of the following model (omitting individual level subscripts):

$$\ln Y = \beta_0 + \gamma STEM + \beta X_1 + \mu \quad , \tag{1}$$

where *ln Y* is the log annual wage, *STEM* is a dummy variable indicating graduation from a STEM subject[9], so that γ is the estimated return to graduating from a STEM field. X_1 are controls for the individual's characteristics, including higher education institution dummies and dummies on employer's postcodes to capture the effect of the local labour market on wages. μ is a random component assumed to be normally distributed. Standard errors are clustered at the three-digit level subject choice to account for possible correlations in error terms by major[10]. To test whether the returns to science are universal or specific to working in a scientific occupation, I extend the base model by including dummies for occupation (O_k) and their interactions with the STEM indicator.

$$\ln Y = \gamma_0 + \gamma_1 STEM + \sum_k \gamma_{2k} O_k + \sum_k \gamma_{3k} O_k STEM + \beta_1 X_1 + \mu_1 \tag{2}$$

As mentioned above, these estimates of the returns to STEM might be biased by selection effects, for example, if more able graduates are the ones studying science

---

9 In alternative specifications, I include a breakdown MIPE other STEM or even a full breakdown by subjects.
10 Results are not sensitive to the choice of cluster. For example, in the favoured specification of log annual earnings (Table3, column 3) the standard errors on STEM are 0.0122, 0.0115, 0.0128 and 0.0159 when clustering at the job location, institution, occupation or subject level, respectively.

and gaining jobs in scientific occupations. I therefore now account for selection in both subject and occupation. To do so, I estimate a double selection model, whereby the decision to do a science degree and work in a scientific occupation are jointly estimated, and are allowed to be correlated. This is an extension of the Heckman (1979) 2-steps selection model, where the first step includes the joint decision to study a STEM topic and to work in a scientific occupation. This joint decision is estimated by a bivariate probit regression. As in Fishe, Trost and Lurie (1981), I analyse the cases where the error terms in the two selection equations are i) uncorrelated and ii) allowed to be correlated. The basic modelling idea stems from a Roy (1951) model whereby workers select the subject/occupation pair (say k_1, k_2 *and* k_3) that maximises their expected earnings ((E(w_j / j = k_l) > (E(w_h / j = k_l)). The model thus consists of two selection equations (graduate from STEM, work in a scientific occupation) estimated simultaneously and three wage equations which are specific to a subject/occupation pair[11]. For each selection equation, I only observe a dichotomous outcome, but this observed outcome results from an unobservable index of the utility of this decision, represented by an upper-script star:

$$\begin{cases} STEM_i^* = \gamma_1 Z_{1i} + \varepsilon_{1i} \\ SOCC_i^* = \gamma_2 Z_{2i} + \varepsilon_{2i} \\ lnW_{ki} = \beta_k X_{ki} + \mu_{ki} \qquad k = 1,2,3 \end{cases} \qquad (3)$$

The two selection equations define three different groups such that:

$$\begin{cases} k = 1 \equiv STEM_i^* \geq 0 \ \& SOCC_i^* \geq 0 \\ k = 2 \equiv STEM_i^* \geq 0 \ \& SOCC_i^* < 0 \\ k = 3 \equiv STEM_i^* < 0 \ \& SOCC_i^* < 0 \end{cases}$$

Unobservable characteristics correlated with wages might also be correlated with either or both selection variables ($E[\mu_{ki}\varepsilon_{ti}] = \sigma_{kt}$); to account for this endogeneity,

[11] I drop the pair non-science degree/scientific occupation due to small sample size and concerns about measurement error defining this group.

the wage regressions can be corrected. Since there are two selection equations, an additional difficulty is to account for the correlation structure between these two selection processes. Two different cases exist, i) the error terms in the two selection equations are independent ($E[\varepsilon_{1i}\varepsilon_{2i}] = 0$) or correlated ($E[\varepsilon_{1i}\varepsilon_{2i}] = \sigma_{12}^{S}$).

For each *k*-type individual, I predict the expected wages in the *k*-type group. Following the notations from Fishe, Trost and Lurie (1981), the expected wage equations in the case of uncorrelated selection equations ($E[\varepsilon_{1i}\varepsilon_{2i}] = \sigma_{12}^{S}$ =0) are defined as follow:

$$\begin{cases} E[lnW_{ki} \,/\, i \in k = 1] = \beta_k X_{ki} - \sigma_{k1} M_{1i} - \sigma_{k2} M_{2i} \\ E[lnW_{ki} \,/\, i \in k = 2] = \beta_k X_{ki} - \sigma_{k1} M_{1i} + \sigma_{k2} M_{4i} \\ E[lnW_{ki} \,/\, i \in k = 3] = \beta_k X_{ki} + \sigma_{k1} M_{3i} + \sigma_{k2} M_{4i} \end{cases} \quad (4.1)$$

The correction terms M_1, M_2, M_3 and M_4 , also known as the inversed Mills ratio, are defined as follows, where the function f(.) and F(.) refers to the density and cumulative normal distributions, respectively:

$$\begin{cases} M_{1i} = f(Z_{1i}\gamma_1)/F(Z_{1i}\gamma_1) \\ M_{2i} = f(Z_{2i}\gamma_2)/F(Z_{2i}\gamma_2) \\ M_{3i} = f(Z_{1i}\gamma_1)/[1 - F(Z_{1i}\gamma_1)] \\ M_{4i} = f(Z_{2i}\gamma_2)/[1 - F(Z_{2i}\gamma_2)] \end{cases} \quad (4.2)$$

When the two selection equations are correlated, the wage equations and selection terms become more cumbersome to compute. Again, following Fishe, Trost and Lurie (1981), they are computed as:

$$\begin{cases} E[lnW_{ki} \,/\, i \in k = 1] = \beta_k X_{ki} + \sigma_{k1} M_{12i} + \sigma_{k2} M_{21i} \\ E[lnW_{ki} \,/\, i \in k = 2] = \beta_k X_{ki} + \sigma_{k1} M_{56i} + \sigma_{k2} M_{65i} \\ E[lnW_{ki} \,/\, i \in k = 3] = \beta_k X_{ki} + \sigma_{k1} M_{78i} + \sigma_{k2} M_{87i} \end{cases} \quad (5.1)$$

Where the selection terms are defined as:

$$M_{lji} = (1 - (\sigma_{12}^{s})^2)^{-1} * \left(P_{li} - \sigma_{12}^{s} P_{ji}\right)$$

$$
\begin{cases}
P_1 = \frac{\int_{-\infty}^{Z_2\gamma_2}\int_{-\infty}^{Z_1\gamma_1} \varepsilon_1 f(\varepsilon_1\varepsilon_2)d\varepsilon_1 d\varepsilon_2}{F(Z_1\gamma_1, Z_2\gamma_2)} \\
P_2 = \frac{\int_{-\infty}^{Z_1\gamma_1}\int_{-\infty}^{Z_2\gamma_2} \varepsilon_2 f(\varepsilon_1\varepsilon_2)d\varepsilon_2 d\varepsilon_1}{F(Z_1\gamma_1, Z_2\gamma_2)} \\
P_5 = \frac{\int_{Z_2\gamma_2}^{\infty}\int_{-\infty}^{Z_1\gamma_1} \varepsilon_1 f(\varepsilon_1\varepsilon_2)d\varepsilon_1 d\varepsilon_2}{F(Z_1\gamma_1, -Z_2\gamma_2)} \\
P_6 = \frac{\int_{-\infty}^{Z_1\gamma_1}\int_{Z_2\gamma_2}^{\infty} \varepsilon_2 f(\varepsilon_1\varepsilon_2)d\varepsilon_2 d\varepsilon_1}{F(Z_1\gamma_1, -Z_2\gamma_2)} \\
P_7 = \frac{\int_{Z_2\gamma_2}^{\infty}\int_{Z_1\gamma_1}^{\infty} \varepsilon_1 f(\varepsilon_1\varepsilon_2)d\varepsilon_1 d\varepsilon_2}{F(-Z_1\gamma_1, -Z_2\gamma_2)} \\
P_8 = \frac{\int_{Z_1\gamma_1}^{\infty}\int_{Z_2\gamma_2}^{\infty} \varepsilon_2 f(\varepsilon_1\varepsilon_2)d\varepsilon_2 d\varepsilon_1}{F(-Z_1\gamma_1, -Z_2\gamma_2)}
\end{cases}
\tag{5.2}
$$

V Results

V.1 – OLS estimates

The first items to discuss are the estimates resulting from the simple OLS model (1) of the average returns to STEM. This is the type of model that has often been estimated in the literature. The top panel of Table 3 reports several specifications of this model. In the simplest specification, controlling only for the local labour market (postcode dummies), I find a return to having graduated from STEM of 11.4%. In specification (2) I control for a rich set of individual characteristics including gender, age at graduation, disability status, but also parental social class, type of school attended and academic ability[12]. Including these controls reduces the science premium by 31% to 7.9%. In column (3) I further control for class of degree and include higher education institution dummies so as to capture the quality/reputation of the education received. This further reduces the return to STEM to 5.8%. Importantly, these last set of controls are often not observable, leading to an upward bias in the returns to science in most of the literature. The returns to studying science appear substantial, similar in scale to the returns to attending the highest quality institutions (Altonji et al., (2015) or Chevalier, 2014). The last two columns report estimates

[12] This is proxied by the A-levels grades. A-levels are the national exams taken at the end of secondary schools in England. A-levels, or their equivalent, determine admission to higher education.

when splitting the population by gender. The returns to studying science are 80% larger for males than for females (8.5% vs 4.8%) which partially comes from differences in subject choice by gender. As seen in Table 1, male scientists are disproportionally found in mathematics, IT, engineering and architecture, majors that have higher mean earnings than psychology and biology, two of the most female-dominated fields (Subject allied to Medicine is also 84% female but has average earnings).

The second panel of Table 3 separates science graduates between MIPE and other STEM. Returns to other STEM are marginally larger in the basic specification but no significant wage differential between MIPE and other STEM is found when the most extensive set of controls is included. Again, there are some differences in returns by gender. For males, returns are very similar when graduating from MIPE or other-STEM, at around 8.5%. Females graduating from other-STEM earn significantly more than non-STEM graduates (+5.3%), and even then, the returns are lower (not significantly) than for males. For females, no significant return to MIPE is found. Below, I explore further whether the differences in returns stem from differences in subject of study or differences in returns by subject.

The last panel of Table 3 reports the wage premiums over a non-science graduate for each of the science majors. This confirms the descriptive statistics that large variations in the returns to science by major exist. Graduates from medical schools earn 51% more than non-STEM graduates. Graduates from engineering, architecture and subjects allied to medicine also enjoy premiums between 10% and 15%, while math and IT graduates' earnings are about 6.5% more than non-STEM graduates. Only psychology graduates have significantly lower earnings than non-STEM majors (-5%). This is a concern when considering that psychology is the discipline, which according to HESA, has seen the largest increase in graduates between 2003 and 2014

(+89%). At this level of disaggregation, there are differences in returns to fields by gender, with male graduates generally enjoying larger returns to a science degree than their female peers (sport science is an exception), but these differences are not statistically significant.

I now test whether the returns to a science degree are universal across occupations or specific to working in a scientific job. This is a crucial test of the argument that science graduates are poached to work in non-scientific occupations. Table 4 reports estimates of the return to STEM when controlling for occupation (using the full specification described in the previous paragraph). When occupation controls are included (column 1), returns to STEM drop to an insignificant 2%. This compares with an occupational premium in scientific, financial and educational occupations of 14% to 16%. In the second column, I add interactions between STEM and occupation groups, to assess whether scientific skills are rewarded in other occupations. None of the interactions are statistically significant. There is no overall return to studying science, and scientific skills are only rewarded in scientific occupations[13]. This conclusion is very similar to Kinsler and Pavan (2015), who estimate the wage returns to science for graduates not working in a job related to their studies to be, as is the case here, an insignificant 2%.

This analysis is repeated in columns (3) and (4), separating the science graduates between MIPE and other STEM. The prior is that MIPE skills are less occupation specific and may generate returns even in non-scientific occupations. This is not supported by the data. The estimated wage returns to MIPE and other STEM are similar and again not statistically significant outside scientific occupations, i.e., there

[13] We also estimate a model excluding individuals who have not studied a STEM subject but still works in a scientific occupation (177 observations), as this may reflect measurement error, but our estimates are unchanged.

is no specific return to being a science graduate of any type in a non-scientific occupation.

In the remaining two columns of Table 4, I report estimates separately by gender. The conclusions are very similar, the wage premiums to studying science is specific to being matched to a scientific occupation. Altogether, the differences between men and women estimates are also never statistically significant.

The science wage premium is match-specific, which is puzzling since one often advocated reason for the leakage of science graduates is that non-scientific occupations value the skills of science graduates and offer higher wages than scientific occupations. This is not supported by the data. The next section assesses whether this conclusion is altered when accounting for selection by subject and occupation.

V.2: Selection model

The econometric section described that the naïve estimates of the returns to science and occupation may be biased if unobservable individual characteristics correlate with subject or occupation, and with earnings. To correct for this selection, I use the double selection model presented in the previous section. This model can only be identified if the set of variables (Z_1 and Z_2) does not fully overlap with X; i.e., exclusion restrictions which determine the choice variables but not earnings are needed. We discuss below the identifying variables.

Due to the presumption that not enough pupils study STEM at tertiary level, grants and other policies reducing tuition fees disproportionally target STEM subjects over other subjects[14]. Indeed, Figure 4 reports local polynomial estimates of the

[14] Two articles have specifically investigated the effect of differential tuition fees on major decision. Stange (2015) relies on time differences between institutions in the introduction of differential pricing of majors, whereby engineer majors pay higher tuition fees. Their introduction leads to a drop in the

fraction of graduates studying STEM by tuition fees status, conditional on ability (A-level score). At all levels of ability, students paying full tuition fees are less likely to be studying STEM. The gap is particularly large for high ability students, where a 15 percentage point gap in the probability of studying STEM exists between those paying full tuition fees and those paying reduced or no fees. Of course, a concern would be that fee status has a direct impact on occupational choice and wages, for example, if needs to repay debt affect sector of work (Field, 2009). This is unlikely to be the case since the level of tuition fees was rather modest (£1,000 per year). Indeed tuition fee status has no statistically significant effect in a wage regression that controls for subject and occupation, nor is it related to occupation. As such, tuition fee status appears to be correlated with subject choice but not directly to subsequent career decisions.

The decision to work in a scientific occupation is identified from intergenerational correlation in occupation; i.e., parents working in a scientific occupation may influence the career decisions of their child (Long and Ferrie, 2013). Figure 5 reports the proportion of graduates in scientific occupation by parental occupation. At all levels of ability, children of scientists are 5 percentage points more likely to be working in a scientific occupation. Again, a concern of this identifying strategy might be that parental occupation allows young graduates to secure higher earnings (nepotism, information,..). Estimating an extension of the previous wage regression, which includes parental occupational, I find no statistical evidence supporting the idea that parental occupation is correlated with child's earnings.

I now estimate the selection model as described in (3), whereby the first step is a bi-variate probit model jointly estimating the decision to study science and to work in

---

share of engineering degrees of 9%. On the contrary, Evans (2013) evaluates the effect of a STEM specific grant on major decisions in Ohio but reports no significant effect.

a scientific occupation.[15]. The two exclusion variables are significantly different from zero individually, and an F-test of their joint significance has a value of 29. The estimated correlation in the residuals of the two selection equations is positive ($\sigma_{12}^{S} = 0.200$), i.e., individuals who have unobserved characteristics making them more likely to study science are also more likely to work in a scientific occupation. However, this correlation is very imprecisely estimated, so I conduct the second step twice, first assuming that the correlation is in fact 0, then using the estimated correlation. The inverse Mills ratio are computed in both cases, and the wage equations for the three groups of interest (*k=1*: STEM in scientific occupation, *k=2:* STEM not in scientific occupation, *k=3:* non-STEM) is estimated. The second panel of Table 5 reports the coefficients on the selection correction terms in the wage equation for each group. Whether correlation of the error terms in the selection equations is assumed or not, the selection terms are only significant for the group of STEM graduates working in scientific occupations. These individuals have unobservable characteristics that are positively correlated with the subject and occupation decisions and their earning ability, maybe due to their greater interest for the subject.

I also report the average predicted wages for the three groups. The results are not dependent on the assumption regarding the correlation of the selection processes. The expected wages of STEM graduates not working in scientific occupations are similar to non-STEM graduates, and the returns to working in a scientific occupation are 20%. Using the observed proportion of STEM graduates in scientific occupation, the average return to a science degree is thus 10.01%.

[15] For this model to converge I exclude postcode and institution dummies. The model was also estimated separately by gender – Convergence was only achieved for the male sub-sample but point estimates were very similar to those obtained with the full population.

Figure 6 reports the distribution of the predicted wage differential in the observed occupation and in the alternative occupations $(ln\widehat{W(k=l)} - ln\widehat{W(k \neq l)}/k = l)$. If individuals choose subjects and occupations maximising their expected earnings, as in a Roy model, this inequality should always be positive, i.e., individuals have greater earnings in their observed subject/occupation than in an alternative. For non-STEM graduates, the expected wage differentials had they studied a STEM subject would have been very different depending on whether they would have worked in a scientific occupation or not (Figure 6A). Had they worked in a scientific occupation, they would have earned close to 100% higher wages. But if following graduation from a STEM subject they would have worked in a non-scientific occupation, they would have been worse off by 40% compared to their realised wages. These estimates are likely to be biased since we do not fully account for subject specific taste but they also reflect the large heterogeneity of expected earnings following graduation from a STEM subject (as is also the case in the US, see Kinsler and Pavan, 2015).

For STEM graduates not working in a scientific occupation (Figure 6B), the distributions of the difference between predicted wages in their observed occupation and in alternative subjects or occupations are almost centred on 0. On average, they would have been 3.7% better off if they had worked in a scientific occupation. As such, there is no strong evidence that this group of graduates has been pulled to work in non-scientific occupations by higher expected wages. They are also marginally better off having studied science and not working in a scientific occupation than if they had not studied science altogether (the mean expected wage differential is 3.7%). If we consider these graduates to be marginal in their choices to study STEM and to work in a scientific occupation, these results could be interpreted as the marginal returns to studying science and to work in a scientific occupation.

Only graduates working in scientific occupations (Figure 6C) mostly conform to a Roy's model of subject/occupational choice. They earn 13% to 18% more than if they did not work in a scientific occupation or had not studied science, respectively, with a clear majority of them having a positive wage differential between their realised wage and their expected alternative wage.

Up to 40% of graduates would appear to have been better off if they had studied another field, but this may have to do with the inability to model taste for specific studies. Indeed, the limited role of future earnings on major choice is highlighted, among others, in Wiswall and Zafar (2015). Using experimental data in which a small group of New York University students was provided with information on the distribution of earnings by major, they conclude that for these high ability students, while the expected probability of graduation and expected earnings are related to field of study chosen, the main determinant is taste for the subject.

Overall, I do not find consistent evidence that students are better off graduating from a STEM subject. For non-STEM graduates, this would have been the case only if they had subsequently worked in a scientific occupation, but not if they had worked in a non-scientific occupation. Across all three groups, I consistently find that graduates would have been (are) better off working in a scientific occupation. For the most marginal group, returns to studying science are low (+3.7%). This does not appear consistent with the often heard hypothesis that non-scientific occupations pull scientists away from science because of higher wages.

VI Pull and Push Factors – Non Financial Concerns

While the results so far have not been consistent with science graduates being pulled away from a career in science by higher wages in other occupations, I now focus on other dimensions of the job that may explain the attraction of non-science occupations for STEM graduates. Table 6 reports OLS estimates of major and occupation on subjective measures reflecting the match between current job and education (panel A) and reasons for having accepted the current job (panel B).

First, I proxy the quality of the job match, relying on the definition of over-education provided in Elias and Purcell (2004). I then estimate a linear probability model similar to the one presented in (2). According to this definition, 26% of non-science graduates are over-educated. Controlling for observable characteristics, a science major reduces the risk of over-education by 22 percentage points, but only for graduates working in scientific occupations. STEM graduates working in non-scientific occupations are 20 percentage points more likely to be over-educated than STEM graduates working in scientific occupations. Since over-education is usually associated with a wage penalty, this is consistent with STEM graduates in scientific occupations earning more than peers working in other occupations.

To further investigate the demand for scientific skills, columns 2 and 3 report estimates on whether the subject of study and skills, respectively, were important in obtaining the current position. For non-scientific graduates, 46% respond that the subject was either a requirement or important in obtaining the job. Again, for science graduates this probability is much larger (+36 pp), but only if working in a scientific occupation. The large majority (85%) of graduates believe that their skills were important in obtaining their current job, but for STEM graduates working in scientific occupation, this is even larger (+4.5pp). Scientific skills do not appear to be in large demand in non-scientific occupations; there is no difference in the self-report between non-STEM and STEM graduates not working in scientific occupation.

The final two columns of Panel A report estimates for overall measures of match quality: satisfaction with career so far and whether with insight the respondent would rather have studied a different subject. One could assume that if science graduates had been attracted to work in non-scientific occupations, they would report higher levels of satisfaction than other graduates. This is not the case. Only science graduates in scientific occupations are more satisfied with their career than non-science graduates, and they report greater levels of satisfactions with their field of study choice.

Over all these outcomes, STEM graduates in scientific occupations have a better match of their skills with their jobs, resulting in greater career satisfaction, less risk of over-education and fewer regrets about field of study than other graduates. These results are not consistent with STEM graduates being pulled to non-scientific occupation. To investigate this issue further, I investigate popular reasons for having accepted the current job (Panel B). We can split these reasons between positive choices—the job fit with my career plan, is exactly the job I wanted, this was the best job offer, or the job allows for broadening skills—and negative reasons such as this was the only job offer, or I took it to pay-off debts. With the exception of "This was the best job offer", STEM graduates working in scientific occupations are more likely than non-STEM graduates to mention one of the positive reasons for being in their current job, and the estimates are always different from the one obtained for STEM graduates not working in a scientific occupation. This later group is mostly indistinguishable from non-STEM graduates, apart from that they are 6 percentage points less likely to be in their current job because it was the "best job offer". Regarding the two negative reasons for being in the current job, all STEM graduates are less likely to be in a job to pay-off debts.

Taken altogether, this evidence does not support the claim that science graduates have been lured to work in non-scientific occupations by better amenities; they are mostly undistinguishable from non-STEM graduates and less positive than STEM graduates working in a scientific occupation.

As robustness checks I estimate three other specifications of these models. First, I include current wage to capture potential effects due to compensating wage differentials, but the conclusions remain broadly unchanged. Second, I estimate these models when adding interactions between gender, and the STEM and occupation status, to assess whether the reasons to work in a given occupation differ by gender. I then conduct an F-test of the joint significance of these interaction terms, reported as F(2,148) Gender in Table 6. The interactions are never significant, indicating that the reasons to be in a job requiring scientific knowledge or not are similar for male and female graduates. Finally, I split the STEM group between MIPE and other-STEM and test whether the coefficients on the interaction between subject group and occupation differ between STEM and MIPE (last two rows of each panel in Table 6). For science graduates not working in a scientific occupation, the only significant difference is that MIPE graduates are more likely to be in a non-graduate job than other STEM graduates. For graduates in scientific occupations, there are more differences between MIPE and other STEM, with other-STEM being better matched than MIPE. Altogether, these results suggest that for science graduates working in a non-scientific occupation, neither the reasons for taking such a position, nor the quality of the match differ between MIPE and other-STEM. More generally, the reasons are not different than those put forward by non-STEM graduates, confirming that science graduates not in scientific occupations, whatever their background, are mostly pushed into these occupations.

VII: Who are the mis-matched STEM?

This section examines whether some observable characteristics of STEM graduates are associated with the probability of not working in a scientific occupation. In particular, since the previous evidence has mostly been consistent with

STEM graduates being pushed to non-scientific occupation, I focus on measures of academic skills: field of study, degree grade, and institution quality, as well as gender. For STEM graduates, I report in Table 7, the estimates from a linear probability model on being observed working in a scientific occupation.

In model 1, I explore the role played by major choice even after controlling for institution fixed effects. Fields of study are correlated with the probability of working in a scientific occupation. Compared to a math graduate, graduates from the more applied science fields—in particular medicine, subjects allied to medicine, engineering and IT—are 30 to 60 percentage points more likely to be observed working in a scientific occupation three years after graduation. Other subjects are largely indistinguishable from math in their probabilities of landing a scientific job. Consistent with selection by ability, class of degree is associated with a greater probability of working in science. There is, in particular, a sharp break at a 2.2 or lower, consistent with most graduate programs and graduate jobs stating that a grade of 2.1 or above is needed to apply. Note that, even after controlling for field of study, males are 9 percentage points more likely than female graduates to be in a scientific occupation.

In the second column, I drop the institution fixed effects and include instead a measures of institution quality[16]. Institutions above median and, in particular, in the top quarter of quality have graduates with a greater probability of working in scientific occupations. Other results are largely unchanged. Graduate characteristics, in particular those correlated with skill accumulations and with more applied skills, are associated with a greater probability of working in a scientific occupation. This is

[16] Quality is measured as the first component in a principal component analysis including research assessment score, student staff ratio, academic expenditures per student, entry grades of students and graduate prospects; measures that are typically included in league tables.

consistent with the push explanations advanced previously for the leakage of scientists; STEM graduates with lower skills are pushed towards non-science jobs.

I then split the sample between MIPE and other STEM (Columns 3 and 4). The selection into scientific job is mostly observed for MIPE graduates for which I found strong effects of grades, institution quality and gender. For other-STEM, none of these factors are correlated with the decision to work in a scientific job. While I previously found that the probability of working in a scientific occupation (Table 1) and its returns were not different between MIPE and other STEM (Table 4), the selection by which STEM graduates end up in a scientific occupation seems to differ, with MIPE selected on observables related to their skills.

Finally, I split the sample by gender. The selection into scientific occupations is correlated with skill-related observables for males, but not for females. This is surprising considering that neither the returns to being matched to a scientific occupation nor the reasons for being in a job significantly differ by gender. This suggests that male and female science graduates opt for different scientific occupations.

Overall, this section concludes on selection by highlighting that STEM graduates working in scientific occupations are positively selected on their skills, especially MIPE and male graduates. This also suggests that the market for scientific occupations is not homogenous.

VIII: Conclusions and comments

A puzzle in the labour market for science graduates is that there is a popular view that there is a shortage of science graduates, but at the same time, 50% of science graduates do not work in scientific occupations. An often advocated reason is that science graduates are pulled to work in non-scientific occupations.

This study provides evidence that the wages of STEM graduates are higher than non-STEM graduates, but this premium is driven by higher wages in scientific occupations, not by a premium for scientific skills across the labour market. This is true for all STEM, but it also holds when splitting STEM between more mathematically oriented subjects (Math, IT, Physics and Engineering) and other STEM. Another contribution of this manuscript is to account for both the selection into field of study and occupation. The self-selection bias in the returns to subject is small—the correction terms are often insignificant—and predicted earnings, when using the selection or a simple model, differ on average by less than 2%. Again, this leads to the conclusion that the returns to scientific skills are specific to being matched to a scientific occupation. Consistent with a sorting of science graduates, the returns to matching for STEM graduates in non-scientific occupations would be low (+ 3.7%).

The career decisions of graduates may also be related to non-financial reasons. However, this study found little support for the notion that non-scientific occupations exert a pulling attraction for science graduates; on the contrary, science graduates in non-scientific occupations are more likely to be over-educated, less likely to report that their subject of study was important to get their job, less likely to be in the job they wanted to do, less satisfied with their careers and less likely to agree that with hindsight they would study the same major. It is thus unclear that the leakage of scientists is due to the appeal of other occupations.

Numerous reports have claimed that there is a shortage of scientists. Indeed, this study estimates a wage premium for working in a scientific occupation, but the other results—such as no returns to STEM skills in non-scientific occupations, lower match quality of STEM graduates in non-scientific occupations and greater regret about field of study among non-matched STEM graduates—caution against the claim that more STEM graduates are needed.

The puzzle of why so many science graduates work in non-scientific occupations remains. It may be due to a mismatch between degree programmes and employers' needs (see the Lambert (2003)), which would also be consistent with the heterogeneity in the returns to science by detailed fields. But then, rather than calling for more scientists to be trained, it would appear that employers should provide the training for graduates to have appropriate job-specific skills that are required. Alternatively, the institutional set-up of education in England, where students specialise early, increases the costs of switching majors. Individuals who become dissatisfied with their major choice are thus trapped and switch only when entering the labour market (Malamud, 2010). Allowing students to specialise later might thus reduce the leakage of scientists (see Bridet and Leighton, 2015, for some simulations of these effects).

Reference:

Altonji, Joseph G., Peter Arcidiacono, and Arnaud Maurel. *The analysis of field choice in college and graduate school: Determinants and wage effects*. No. w21655. National Bureau of Economic Research, 2015.

Arcidiacono, Peter. "Ability sorting and the returns to college major." *Journal of Econometrics* 121, no. 1 (2004): 343-375.

Berger, Mark C. "Predicted future earnings and choice of college major." *Industrial & Labor Relations Review* 41, no. 3 (1988): 418-429.

Bridet, Luc, and Margaret Alice Leighton. "The major decision: Labor market implications of the timing of specialization in college." University of St Andrews, mimeo (2015).

Cappelli, Peter. "Skill Gaps, Skill Shortages and Skill Mismatches: Evidence for the US." *NBER Working Paper Series* (2014): 20382.

Chevalier, Arnaud. "Subject choice and earnings of UK graduates." *Economics of Education Review* 30, no. 6 (2011): 1187-1201.

Chevalier, Arnaud. "Does higher education quality matter in the UK." *Research in Labor Economics* 40 (2014): 257-292.

Council of Canadian Academies (2015) “Some Assembly Required: STEM Skills and Canada’s Economic Productivity” http://www.scienceadvice.ca/uploads/ENG/AssessmentsPublicationsNewsReleases/STEM/STEMFullReportEn.pdf

Elias, Peter and Kate Purcell (2004), “SOC(HE): A classification of occupations for studying the graduate labour market”, *ESRU*, research paper 6

European Parliament. (2015) “Encouraging STEM Studies for the Labour Market”. Directorate General for Internal Policies **IP/A/EMPL/2014-13.**

Evans, Brent J. "SMART Money: Do Financial Incentives Encourage College Students to Study Science?." *Society for Research on Educational Effectiveness* (2012).

Field, Erica. "Educational Debt Burden and Career Choice: Evidence from a Financial Aid Experiment at NYU Law School." *American Economic Journal. Applied Economics* 1, no. 1 (2009): 1.

Fishe, Raymond PH, Robert P. Trost, and Philip M. Lurie. "Labor force earnings and college choice of young women: an examination of selectivity bias and comparative advantage." *Economics of Education Review* 1, no. 2 (1981): 169-191.

Hamermesh, Daniel S., and Stephen G. Donald. "The effect of college curriculum on earnings: An affinity identifier for non-ignorable non-response bias." *Journal of Econometrics* 144, no. 2 (2008): 479-491.

Hastings, Justine S., Christopher A. Neilson, and Seth D. Zimmerman. *Are some degrees worth more than others? Evidence from college admission cutoffs in Chile*. No. w19241. National Bureau of Economic Research, 2013.

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

Ketel, Nadine, Edwin Leuven, Hessel Oosterbeek and Bas Van der Klaauw. “The returns to medical school: Evidence from admission lotteries”. *American Economic Journal: Applied Economics*. (Forthcoming) 2016.

Kinsler, Josh and Ronni, Pavan. "The Specificity of General Human Capital: Evidence from College Major Choice." Journal of Labor Economics,33 (4), (2015), 933-972.

Kirkebøen, Lars, Edwin Leuven, and Magne Mogstad. *Field of study, earnings, and self-selection*. Journal of Political Economy (forthcoming) 2016.

Lambert R. (2003) “Review of Business-University Collaboration” HMSO: London.

Long, Jason, and Joseph Ferrie. "Intergenerational occupational mobility in Great Britain and the United States since 1850." *The American Economic Review* 103, no. 4 (2013): 1109-1137.

Malamud, Ofer. "Breadth versus depth: the timing of specialization in higher education." *Labour,* 24, no. 4 (2010): 359-390.

President's Council of Advisors on Science and Technology (2012) "Engage to Excel" https://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-engage-to-excel-final_2-25-12.pdf

Roberts G. (2002) "SET for Success", HMSO: London.

Roy, Andrew Donald. "Some thoughts on the distribution of earnings." *Oxford economic papers* 3, no. 2 (1951): 135-146.

Sainsbury D. (2006) "The Race to the Top", HMSO: London

Stange, Kevin. "Differential pricing in undergraduate education: effects on degree production by field." *Journal of Policy Analysis and Management*, 34, no. 1 (2015): 107-135.

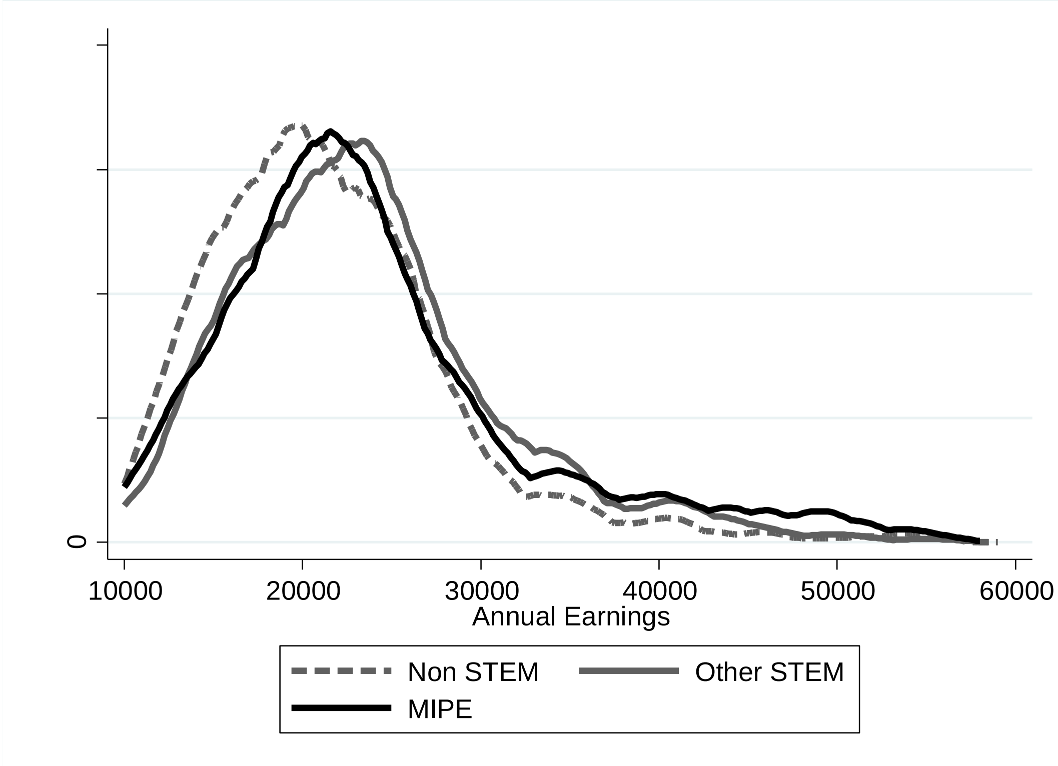
Tipping S. and R. Taylor (2007) "Destination of Leavers from Higher Education Longitudinal Survey 2002/3 Cohort: Assessment of robustness and fitness for purpose" *Higher Education Statistical Agency*.

UK Commission for Employment and Skills (2013). "The Supply and Demand for High Level STEM Skills". https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/302969/high-level-stem-skills-exec-summary-full.pdf

Walker, Ian, and Yu Zhu. "Differences by degree: Evidence of the net financial rates of return to undergraduate study for England and Wales." *Economics of Education Review* 30, no. 6 (2011): 1177-1186.

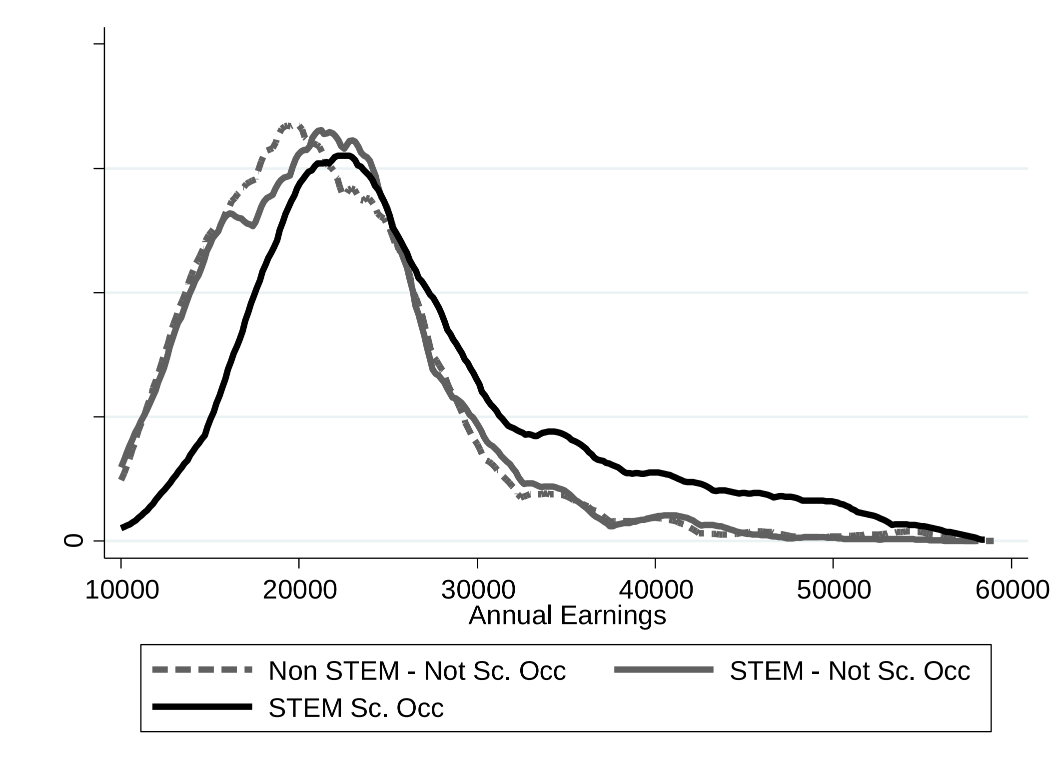
Wiswall, Matthew, and Basit Zafar. "Determinants of college major choice: Identification using an information experiment." *The Review of Economic Studies* 82, no. 2 (2015): 791-824.

Figure 1: Distribution of Annual Earning by Field of Study (October 2006)



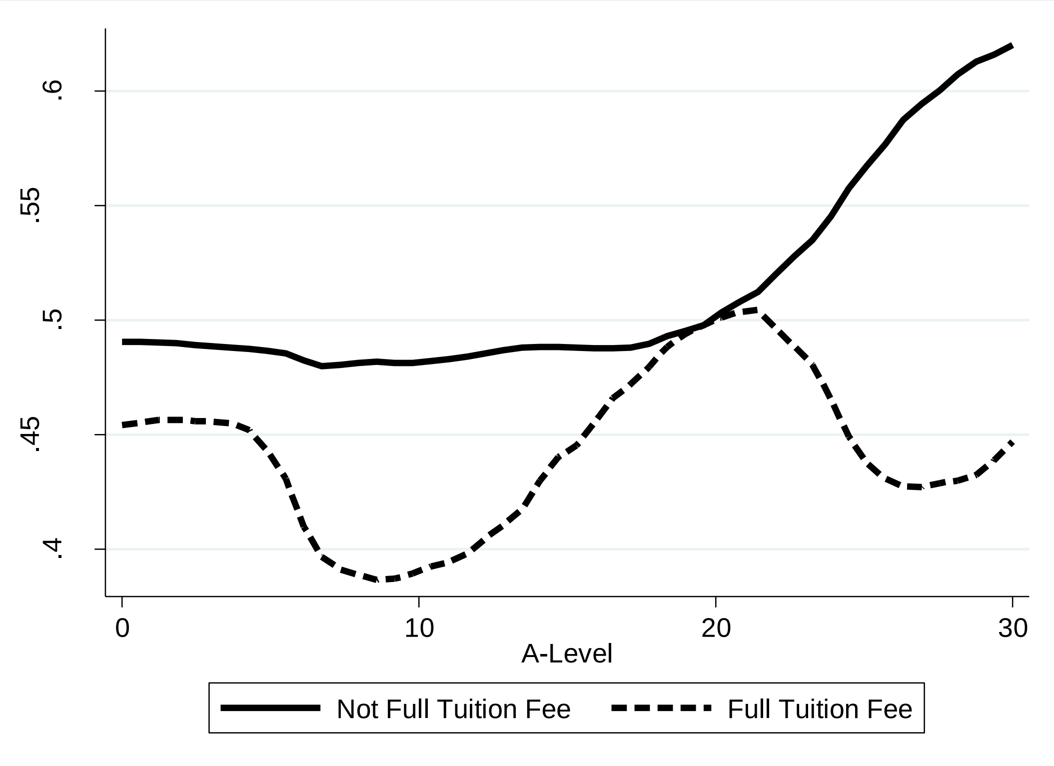
Note: Source LDLHE, Full time employees only- maximum annual earnings trimmed at £60,000. Epanechnikov kernel density.
MIPE: Math, Computing science, Physics, Engineering and Technology
Other STEM: Medicine, Subject allied to Medicine, Biology, Veterinary, Agriculture, Architecture

Figure 2: Distribution of Annual Earning by STEM Status and Occupation Type (October 2006)



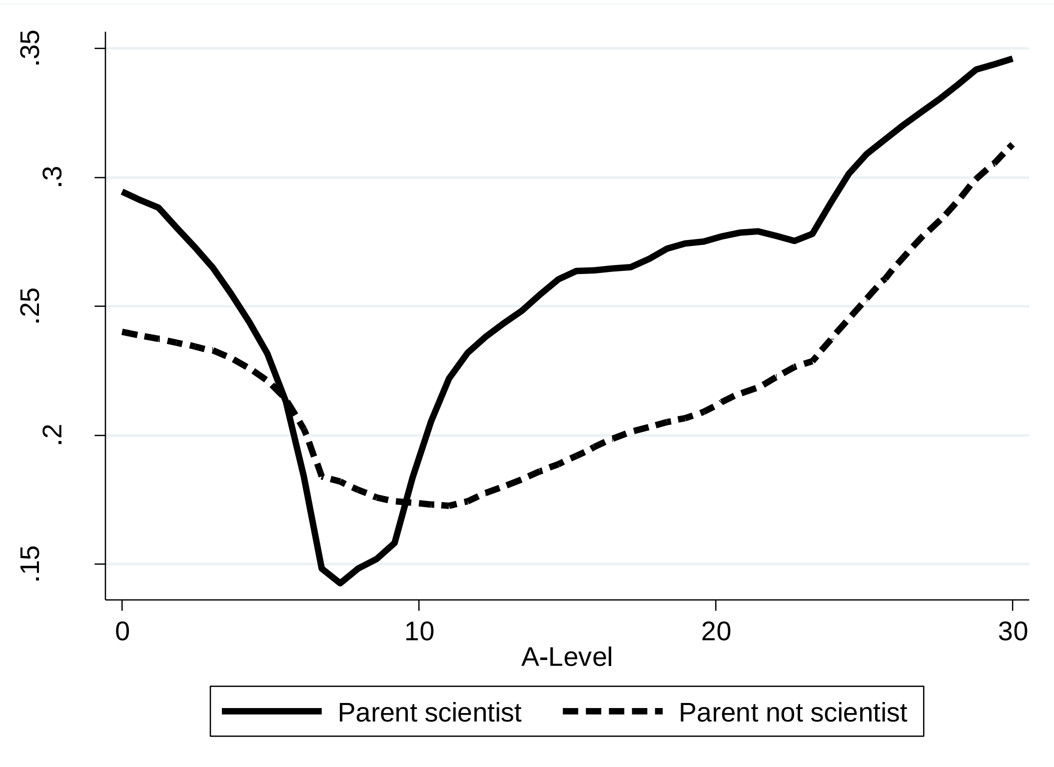
Note: Source LDLHE, Full time employees only- maximum annual earnings trimmed at £60,000. Epanechnikov kernel density. Scientific occupation (Sc. Occ) defined as in Table 1.

Figure 3: Fraction Studying STEM by Tuition Fee Status and Academic Ability



Note: Source LDLHE . Local polynomial estimates based on Epanechnikov kernel

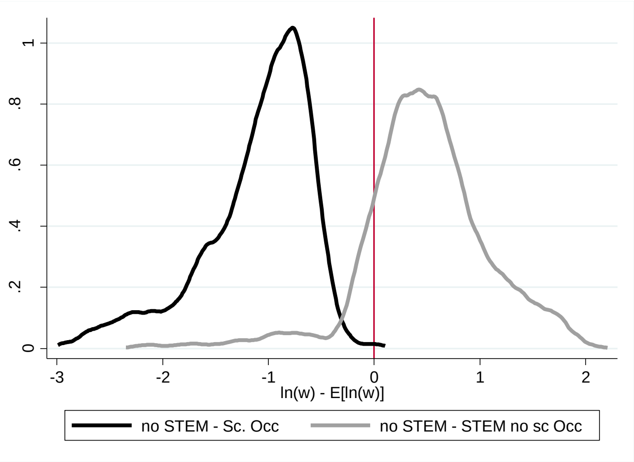
Figure 4: Fraction working in scientific occupation by grade and parental occupational choice



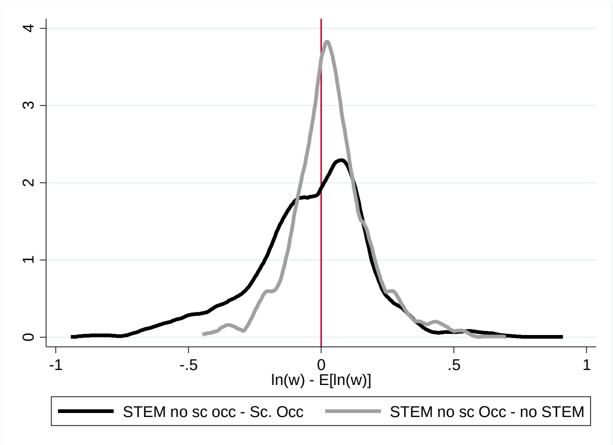
Note: Source LDLHE. Local polynomial estimates based on Epanechnikov kernel

Figure 5: Predicted Earnings Differential in Observed Occupation and Alternative Occupations, by Observed Subject Choice

A] Non STEM graduates



B] STEM graduates, not in sc. Occupation



C] STEM graduates in sc. Occupation

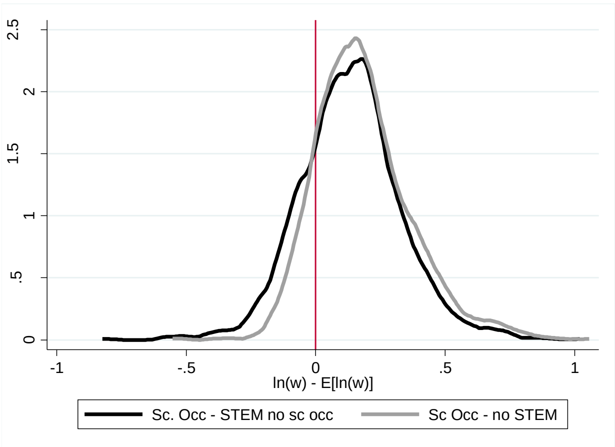


Table 1 Proportion of graduates working in specific occupational group and gender composition

Subject	Scientific occupation	Financial occupation	Education sector	Other	% Male	Obs.
Science subject:						
Medicine and Dentistry	0.95	0.01	0.00	0.04	41.8	390
Sub. allied to Medicine	0.80	0.01	0.03	0.16	16.8	944
Biology, vet, agriculture	0.30	0.01	0.14	0.55	30.3	378
Psychology	0.23	0.02	0.20	0.55	14.1	305
Sport Sciences	0.01	0.04	0.31	0.65	53.0	112
Physical science	0.30	0.04	0.14	0.52	59.1	332
Mathematics	0.25	0.20	0.19	0.46	53.8	203
IT	0.47	0.04	0.06	0.043	75.1	630
Engineering and Tech.	0.59	0.03	0.03	0.35	85.3	575
Architecture & Planning	0.53	0.00	0.00	0.47	76.5	183
Mixed 100% science	0.43	0.14	0.06	0.37	57.5	120
Aggregated subjects						
Non-science	0.05	0.07	0.17	0.72	37.5	4525
STEM (all)	0.43^{+}	0.04^{+}	0.11^{+}	0.42^{+}	49.7	4851
MIPE	0.45^{+}	0.06^{+}	0.08^{+}	0.43^{+}	72.3	1740
Total	0.24	0.05	0.14	0.57	43.5	9376

Note: Observations weighted to be nationally representative

$^{+}$ denotes that the mean is statistically different from the mean for the non-scientific graduates

Science occupations are defined as the following SOC2000 codes: Managers in construction (1122), mining and energy (1123), IT (1136), R&D (1137), Health services (1181), Pharmacy (1182) Healthcare practise (1183), Farm (1211), Natural environment (1212), Chemist (2111), Biologist (2112), Physicists/mathematicians (2113), Engineer (2121. 2122, 2123, 2124. 2125, 2126, 2127, 2128, 2129), IT professional (2131), software professional (2132), medical occupation (2211), other medical professionals (2212), Pharmacist (2213), Optician (2214), Dentist (2215), Veterinarian (2216), Scientific researcher (2321), statisticians (24234), Actuaries (24235), Architects (24310), Technician (3111, 3112, 3113, 3114, 3115, 3119, 3121), draughtsperson (3122), building inspector (3123), IT technician (3131), Nurse (3211), Midwife (3212), Paramedic (3213), other medical associate professional (3214,3215, 3216, 3217,3218, 3221, 3222, 3223, 32290, 32291, 32292, 32293).

Financial occupations are defined as: Financial institution manager (1151), Chartered and certified accountant (2421), Management accountant (2422), Management consultants, actuaries, economists and statisticians (2423), finance and investment analyst (3534), taxation expert (3535), financial and accounting technicians (3537).

Education sector is defined as all occupation in the group teaching professionals (231)

Table 2: Average Annual Earnings by subject of study and occupation

Subject	Mean Earning	Mean Earning and works in Science	Mean Earning and works in Finance	Mean Earning and works in Teaching	Mean earning and other occupations
Science Subjects					
Medicine and Dentistry	39,133	38,909			
Sub. allied to Medicine	24,580	25,074*		21,581	22,948
Biology, vet, agriculture	20,294	20,217		20,822	20,178
Physical science	21,612	22,079		23,226	20,649
Mathematics	24,693	30,432*	27,334	22,802	22,162
Engineering and Tech.	24,934	26,058*		20,308	22,592
Architecture and Planning	24,476	25,150			23,812
Sport science	20,552			20,938	20,207
Psychology	19,285	18,924		19,310	19,355
IT	22,792	24,618*	22,761	23,248	20,712
Mixed 100% science	22,436	23,043			20,825
Aggregated Subjects					
Non science	21,600	22,028	25,854	22,577	20,939
Other-STEM	23,757	26,390*	26,583	22,039	21,197
MIPE	23,488	25,190*	26,740	22,735	21,399
Total	22,677	25,979	26,125	22,352	21,032

Note: Source LDLHE 02/03. Weighted to be nationally representative. Sample restricted to Full time employees with annual salaries lower than £60,000. – means for cells with less than 20 observations are not reported.

* indicates significant difference (95% confidence level) between earning in scientific occupation and all other occupations

Table 3: OLS – (log)Annual Earnings by Field of Study

Panel A	All Base (1)	All Pre-uni controls (2)	All Graduation controls (3)	Male (3)	Female (3)
STEM	0.114**	0.079**	0.058**	0.085**	0.048**
	[0.010]	[0.010]	[0.016]	[0.024]	[0.018]
Panel B					
MIPE	0.104**	0.071**	0.057**	0.082**	0.030
	[0.013]	[0.014]	[0.015]	[0.023]	[0.020]
Other STEM	0.121**	0.084**	0.060**	0.088**	0.053**
	[0.012]	[0.012]	[0.023]	[0.038]	[0.023]
Panel C					
Medicine	0.677**	0.533**	0.515**	0.564**	0.453**
	[0.023]	[0.028]	[0.047]	[0.068]	[0.052]
Subject allied to Medicine	0.161**	0.150**	0.141**	0.238**	0.117**
	[0.016]	[0.015]	[0.034]	[0.056]	[0.032]
Biology, Veterinary	-0.045*	-0.047*	-0.043	-0.006	-0.061*
	[0.027]	[0.025]	[0.028]	[0.044]	[0.034]
Physical science	0.039*	0.018	0.015	0.048**	0.008
	[0.023]	[0.024]	[0.017]	[0.024]	[0.028]
Mathematics	0.105**	0.068**	0.063**	0.054	0.055
	[0.031]	[0.028]	[0.024]	[0.049]	[0.024]
Engineering and Techno.	0.164**	0.127**	0.100**	0.120**	0.073
	[0.019]	[0.020]	[0.025]	[0.028]	[0.053]
Architecture and Planning	0.162**	0.154**	0.134*	0.143	0.144**
	[0.042]	[0.041]	[0.071]	[0.093]	[0.044]
Sport sciences	0.023	0.035	0.024	0.027	0.094**
	[0.042]	[0.043]	[0.018]	[0.031]	[0.027]
Psychology	-0.062**	-0.061**	-0.052**	-0.040	-0.059**
	[0.026]	[0.026]	[0.013]	[0.037]	[0.015]
IT	0.068**	0.068**	0.065**	0.111**	-0.004
	[0.023]	[0.024]	[0.014]	[0.024]	[0.029]
Mixed 100% science	0.052	0.020	0.005	0.034	-0.013
	[0.051]	[0.049]	[0.019]	[0.029]	[0.030]
Socio-economic		Yes	Yes	Yes	Yes
HE controls			Yes	Yes	Yes

Note: N=8280, reweighted to be nationally representative. Standard errors are adjusted for clustering at the subject level (150 clusters). ** indicates statistical significance at the 95% confidence interval.
The omitted subject category is all non-science degree.
STEM indicates all science subjects. MIPE indicates graduates from Math, IT, Physics or Engineering programs
(1) includes a set of dummies for postcode of employer (3 digit)
(2): (1) + controls for A-levels score, a dummy for missing A-levels score, a dummy for female, a set of dummy for parental social class, ethnicity, age on graduation, disability status, and type of previous institution attended.
(3): (2) + dummies for class of degree and institution dummy
.

Table 4: OLS – (log) Annual Earnings by STEM and Occupation Type

	All (1)	All (2)	All (3)	All (4)	Male (4)	Female (4)
STEM	0.021	0.024	0.024	0.023	0.063*	0.012
	[0.015]	[0.017]	[0.020]	[0.021]	[0.036]	[0.023]
MIPE			0.017	0.024	0.032	0.021
			[0.017]	[0.021]	[0.030]	[0.028]
Scien. occ	0.146**	0.105**	0.146**	0.105**	0.120**	0.116**
	[0.014]	[0.021]	[0.014]	[0.021]	[0.040]	[0.032]
Scien. Occ * STEM		0.045		0.053	0.044	0.036
		[0.027]		[0.034]	[0.059]	[0.039]
Scien. Occ * MIPE				0.036	0.029	0.031
				[0.030]	[0.047]	[0.055]
Finance	0.137**	0.145**	0.137**	0.145**	0.144**	0.141**
	[0.018]	[0.024]	[0.018]	[0.024]	[0.032]	[0.038]
Finance * STEM		-0.019		0.013	0.005	0.051
		[0.027]		[0.034]	[0.064]	[0.057]
Finance * MIPE				-0.039	-0.009	-0.093
				[0.034]	[0.048]	[0.061]
Teaching	0.156**	0.170**	0.156**	0.170**	0.134**	0.200**
	[0.020]	[0.021]	[0.020]	[0.021]	[0.035]	[0.024]
Teaching * STEM		-0.034		-0.039	-0.118	-0.017
		[0.028]		[0.039]	[0.079]	[0.037]
Teaching * MIPE				-0.023	0.017	-0.061
				[0.036]	[0.049]	[0.043]
Socio-economic	Yes	Yes	Yes	Yes	Yes	Yes
HE controls	Yes	Yes	Yes	Yes	Yes	Yes

Note: N=8280, reweighted to be nationally representative. Standard errors are adjusted for clustering at the subject level (150 clusters). ** indicates statistical significance at the 95% confidence interval.
The omitted subject category is all non-science degree.
In column 1 and 2, STEM indicates all science subjects. In column 3 to 4 MIPE indicates graduates from Math, IT, Physics or Engineering programs, STEM then indicates graduates from other science programs.

Table 5: Double Selection Model – (log) Annual Earnings, STEM and Occupation

1st step	Selection to STEM	Selection to Science. occupation	ln Wage not STEM	ln Wage STEM not in Science occupation	ln Wage STEM in Science occupation
Paid Full Fee	-0.218**				
	[0.080]				
Parent in science occ.		0.511**			
		[0.186]			
	σ_{12}^{S} = 0.200				
	[0.396]				
Chi(2)	15.63				
2nd Step					
A] σ_{12}= 0					
IMR1			-0.004	0.181*	0.085
			[0.147]	[0.104]	[0.129]
IMR2			0.023	-0.045	0.325**
			[0.046]	[0.078]	[0.144]
E[ln(wage)]			9.91	9.92	10.13
B] σ_{12}~=0					
IMR1			0.015	0.034	0.401**
			[0.056]	[0.026]	[0.169]
IMR2			-0.030	0.007	0.466**
			[0.057]	[0.007]	[0.168]
E[ln(wage)]			9.91	9.92	10.12
Observation	8103		3800	2306	1997

Note: Standard error obtained from bootstrap (500 reps) allowing correlation at the subject level (block bootstrap).
Other controls include gender, quadratic in A-level score, disability status, race dummy and school type dummy.

Table 6: Push and Pull factors by science major and occupation -

Panel A	Non-graduate job	Subject important to get job	Skills important to get job	Satisfied with career	Would choose # subject
1- Non-STEM	{0.258}	{0.458}	{0.850}	{0.838}	{0.357}
2- STEM, not in sc. Occ	0.028 [0.022]	-0.006 [0.031]	0.006 [0.018]	-0.006 [0.016]	0.043* [0.025]
3- STEM, in sc. occ	-0.218** [0.016]	0.357** [0.027]	0.045** [0.016]	0.076** [0.018]	-0.093** [0.032]
F(1,149) (2=3)	116.23**	198.65**	3.99**	16.24**	13.23**
F(2,148) Gender	1.06	0.20	0.67	2.23	0.55
F(1,149) MIPE = other STEM, not in sc. occ	4.10**	0.71	0.86	1.60	0.13
F(1,149) MIPE = other STEM, in sc. occ	0.17	3.69*	1.84	0.00	0.58

Table 6 to be continued

Table 6 continues

Panel B	Job fitted with career plan	Job I wanted	Best Job Offer	To broaden skills	Only Job Offer	Pay off debts
1- Non-STEM	{0.640}	{0.497}	{0.459}	{0.622}	{0.184}	{0.281}
2- STEM, not in sc. Occ	-0.30 (0.022)	-0.027 [0.022]	-0.060** [0.018]	0.001 [0.018]	-0.029 [0.018]	-0.065** [0.018]
3- STEM, in sc. occ	0.138** (0.028)	0.108** [0.033]	-0.002 [0.024]	0.059** [0.023]	0.005 [0.021]	-0.055* [0.020]
F(1,149) (2=3)	31.71**	16.86**	6.02**	8.08**	2.02	0.13
F(2,148) Gender	0.15	0.05	2.29	0.79	0.12	1.62
F(1,149) MIPE = other STEM, not in sc. Occ	0.28	0.20	0.02	2.21	1.27	0.01
F(1,149) MIPE = other STEM, in sc. occ	3.64*	6.83***	2.58	1.83	2.88*	0.35

Note: Standard errors are adjusted for clustering at the 3-digit subject level (150 clusters). * and ** indicates statistical significance at the 90% and 95% confidence interval respectively. Mean values of dependent variable for non-STEM graduates reported in {}. The analysis is based on specification (3) details of which can be found in the note under Table 4.
"Non-graduate job" is defined using Elias and Purcell (2004) which defines 5 categories of graduate jobs 1 Traditional occupation, 2 Modern occupation, 3 New occupation, 4 Niche occupation, 5 Non-graduate job.
"Subject (skills) important to get job" is recoded into a dummy for respondents replying "a formal requirement" or "important".
"Career satisfaction" is coded as 1 for respondents who are very satisfied or fairly satisfied, and 0 for everybody else.
"Would study the same subject include 4 categories": 1 very likely different , 2 likely different , 3 not likely different, 4 not likely at all different, I recode the first two categories as 1 and the remaining two as 0.
Variables in Panel B are answers to reasons for choosing current jobs – all reasons that apply are coded as 1.
F(1,149) (2=3) reports an F-test of whether the coefficient for STEM, not in scientific occupation is significantly different from the coefficient on STEM, in scientific occupation
F(2,148) Gender, is an F-test on the joint significance of interactions between gender and STEM, not in scientific occupations and gender and STEM in scientific occupation. The model is not reported here.
F(1,149) MIPE=Other STEM, not in sc.occ and F(1,149) MIPE=Other STEM, in sc.occ are F-test on the joint the equality of the coefficients on MIPE and other STEM for graduates in non-scientific and scientific occupations respectively

Table 7: Linear Probability Model: Observed in Scientific Occupation

	All STEM	All STEM	MIPE	Other-STEM	Female	Male
Male	0.089**	0.087**	0.201***	0.004		
	(0.036)	(0.034)	(0.051)	(0.031)		
Class degree:2.1	-0.045	-0.066**	-0.073	-0.036	-0.036	-0.076*
	(0.036)	(0.032)	(0.059)	(0.028)	(0.037)	(0.045)
Class degree 2.2	-0.079**	-0.103***	-0.173***	-0.016	-0.022	-0.154***
	(0.035)	(0.035)	(0.049)	(0.035)	(0.036)	(0.051)
Class degree Pass	-0.145**	-0.182***	-0.211***	-0.010*	-0.069	-0.232***
	(0.044)	(0.041)	(0.060)	(0.050)	(0.063)	(0.051)
Unclassified degree	-0.044	-0.044	-0.137*	0.053	0.097*	-0.152*
	(0.053)	(0.060)	(0.077)	(0.066)	(0.049)	(0.079)
Institution Q2		0.011	0.018	0.007	0.015	0.000
		(0.030)	(0.043)	(0.041)	(0.040)	(0.042)
Institution Q3		0.045*	0.124**	-0.008	0.018	0.072*
		(0.026)	(0.054)	(0.029)	(0.034)	(0.037)
Institution Q4		0.064**	0.141**	0.024	0.049	0.068
		(0.029)	(0.055)	(0.029)	(0.035)	(0.048)
Medicine	0.636***	0.629***		Omitted	0.632***	0.623***
	(0.094)	(0.093)			(0.061)	(0.134)
Subject allied to Medicine	0.578***	0.0607***		0.033	0.725***	0.400***
	(0.087)	(0.088)		(0.059)	(0.051)	(0.134)
Biology/veterinary/ agriculture	0.147*	0.124		-0.441***	0.236***	-0.026
	(0.078)	(0.082)		(0.059)	(0.043)	(0.153)
Physics	0.097	0.087	0.076		0.165***	0.039
	(0.082)	(0.084)	(0.068)		(0.051)	(0.130)
Engineering/ Technoloy	0.319***	0.345***	0.298***		0.330***	0.302**
	(0.093)	(0.093)	(0.086)		(0.209)	(0.142)
Architecture/Planning	0.263	0.280		-0.247	0.335	0.230
	(0.181)	(0.204)		(0.192)	(0.209)	(0.240)
Sport science	-0.134*	-0.156**		-0.710***	-0.032	-0.237*
	(0.076)	(0.074)		(0.058)	(0.034)	(0.125)
Psychology	0.073	0.053		-0.522***	0.171***	-0.139
	(0.070)	(0.073)		(0.051)	(0.032)	(0.121)
IT	0.268***	0.264***	0.231***		0.169***	0.270**
	(0.074)	(0.077)	(0.066)		(0.047)	(0.128)
Mixed science	0.192**	0.168**		-0.392***	0.218***	0.145
	(0.078)	(0.073)		(0.054)	(0.034)	(0.125)
Institution fixed effects	Yes	No	No	No	No	No
N	4851	4466	1647	2819	2514	1952

Note: Standard errors adjusted for clustering at the subject level.
Other controls include gender, age dummies, quadratic in A-level score, disability status, race dummy, school type dummy, parental social class and accommodation type while studying. Math is the omitted subject category.

Appendix:

Table A1: Sample Selection:

Selection criteria	Number of observations
Original sample	19,979
First degree only	11,866
Age on graduation [19,25]	9,850
Not special entry student	9,738
Employed FT or PT	9,296

Table A2: Occupational choice of science graduates 6 months and 3 years after graduation.

		Occupation: 3 years after graduation					
		Scientific	Finance	Teaching	Other	Total	Obs.
Occupation: 6 months after graduation	Scientific	[84%] (63%) 1,322	[1%] (9%) 14	[1%] (2%) 12	[14%] (11%) 222	(32%)	1,570
	Finance	[8%] (0%) 8	[53%] (30%) 50	[7%] (1%) 6	[32%] (1%) 30	(2%)	94
	Teaching	[7%] (1%) 12	[0%] (1%) 1	[73%] (24%) 129	[19%] (2%) 34	(4%)	176
	Other	[22%] (18%) 373	[3%] (32%) 53	[9%] (29%) 155	[66%] (55%) 1,130	(35%)	1,711
	Not working	[30%] (19%) 391	[4%] (28%) 46	[18%] (44%) 241	[48%] (31%) 622	(27%)	1,300
	Total Observation	[43%] 2,106	[3%] 164	[11%] 543	[42%] 2,038		4,851

Note: In each cell the percentage in brackets pertains to the row percentage, the percentage in parentheses reports the column's percentage; the last row is the number of observations in the cell.