

DOES HIGHER EDUCATION QUALITY MATTER IN THE UK?

Arnaud Chevalier*

Preliminary: please do not quote

Version 3.0 : July 2014

Abstract:

This paper estimates the financial returns to higher education quality in the UK. To account for the selectivity of students to institution, we rely on a selection on observable assumptions. We use several estimates including the Generalised Propensity Score of Hirano and Imbens, which relies on a continuous measure of institutional quality. This highlights that the returns to quality are heterogenous, and mostly driven by high quality institutions. Moving from an institution in the 3rd quality quartile to a top quality institution is associated with a 7% increase in earnings.

Key words: College Quality, Returns to Education, Generalised Propensity Score
JEL Codes: I22, J31

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 (RES-173-27-0040). I thank Stijn Broecke, Victor Lavy, Tarja Viitanen and participants at seminars at DIUS and Oxford University for comments on earlier drafts, as well as three anonymous referees and Kostas Tatsiramos for detailed comments that have substantially improved the manuscript.

* Royal Holloway University of London, and IZA. Contact: arnaud.chevalier@rhul.ac.uk. Chevalier is also affiliated with CEE (LSE) and SFI (Copenhagen)

I Introduction

Reports of high financial returns to higher education are often trumpeted by stake holders to promote attending higher education. A considerable amount of empirical evidence supports this claim but there are far less evidence on the variability of this return¹, and how it is affected by the quality of the educational input. In general, sorting whereby more able students are selected by more selective institutions and earn more in the labour market (Hoxby, 2009) generates a positive correlation between institutional quality and future earnings. Thus naïve estimates, assuming random allocation of students to institutions, are severely upward biased. Whether returns differ by institution is crucial to inform prospective students about to invest in tertiary education. This paper provides estimates of the effect of institutional quality on early career earnings in the UK using several identification strategies, including, for the first time in this context, the Generalised Propensity Score (Hirano and Imbens, 2004).

While graduates from higher quality institutions earn more (see James et al, 1989 for early evidence) whether this is due to the quality of the institution remains debatable. Assuming that the better inputs improve educational attainment², the human capital model predicts that graduates from higher quality institutions will be more productive and obtain higher wages. In fact, even in the absence of increased human capital, graduates may obtain higher wages if these institutions endow them with a set of peers that improve their job prospects. However, there are reasons to believe that the positive correlation between institution quality and graduate earnings

¹ The Economist (05th April 2014) suggests that some US colleges have even negative financial returns.

² Bowman and Mehay (2002) find a positive effect of attending private and higher-rated institutions on appraisal and promotion, consistent with the earnings effects estimated elsewhere. Robst (1995) also reports that graduates from higher ranked institutions are less likely to experience over-education. Smith (2013) using twins also reports that quality reduces drop-out.

is not causal. As in a signalling model, students may attend more selective institutions to signal their greater ability to future employers³.

Obtaining unbiased estimates of the effect of higher education quality on earnings is difficult since students and institutions select each other, and the characteristics, observable or not, which affect these choices are also correlated with labour market outcomes. To account for selection, researchers have relied on adding controls for students' ability (Brewer et al. 1999), matching models (Black and Smith, 2004, 2006), instrumental variable (Long, 2008), twins/family fixed effects (Behrman et al., 1996; Lindahl and Regnér, 2009), regression discontinuity (Hoekstra, 2009; Saavedra, 2009) or pairing students applying to the same institution (Dale and Krueger, 2002, 2014; Broecke, 2012). Here, we rely on different strategies assuming selection on observables, including propensity score matching and Generalised Propensity Score (GPS); i.e. we assume that the observed match of a student to an institution is based on observable characteristics. Part of the selection process is unknown to us: whether to attend tertiary education, which institution to apply to. However, conditional on applying, the centralised application system in use in the UK makes the assumption of selection on observable characteristics of the prospective student plausible. The UK is characterised by a centralised application system, whereby all prospective students complete the same standardized form, and institutions base their decision to accept or not a student only on this information⁴. In the data at our disposal, we have (almost) the same information than the admission

³ Distinguishing between the mechanisms is relevant for policy. Only if the premium is generated by an increase in human capital does the economy gain from institutional quality, and there is then some possible justification for subsidizing students to attend higher quality institution. If the returns are only due to signalling or networking, then the returns to quality are only private. The data available to us do not offer opportunities to test these hypotheses convincingly Hershbein (2013) proposes a test of the origin of the returns to attend more selective institutions and argues that selectivity provides a positive signal to employers. Jawagushi and Ma (2008) provide some evidence supporting human capital

⁴ There are a few exceptions to this general rule, which we explain later

officers, which makes the assumption of selection on observable plausible⁵. Compare to a propensity score matching estimate, GPS allows us to capture the heterogeneity in the returns to quality, by estimating the effect of educational quality at all points of the quality distribution.

The debate on the effect of institutions on earnings has received a large amount of attention in the U.S. where the market for higher education is fiercely competitive, but much less in the rest of the world. There are a few empirical evidences for the U.K. but their results are somehow ambiguous. To inform this debate, we use survey data pertaining to the 2003 cohorts of graduates, which is linked to administrative data so that background information on the students and academic performance can be added. Another difficulty in this literature is to measure quality as there is no agreement on which input matters. We measure institutional quality as the principal component of a set of education inputs: Research Assessment score, student/staff ratio, academic expenditures per student, mean entry grade and graduate prospect, which are commonly used to create league tables.

We demonstrate that there is considerable heterogeneity in returns to quality, with almost no returns for below median quality institutions and large returns for attending the most prestigious institutions. This is consistent with theoretical predictions that as attendance to college increases, returns to quality increase (Hoxby, 2009). Due to this strong non-linearity, OLS estimates are biased upwards and considerably larger than those obtained by GPS. Defining quality as a discrete step function somehow helps to capture the non-linearity. The larger impact of higher education quality at the high end of the distribution is also present in quantile

⁵ Some institutions, typically Oxford and Cambridge universities, as well as medical, dentistry and veterinary schools also rely on an interview to assess the suitability of the candidates. We conduct robustness checks excluding these graduates.

regressions, thus graduates with the higher earning potentials are the one who benefit the most from attending a high quality institutions.

The rest of the paper is organised as follow. Section 2 reviews the U.S. and non-U.S. literature on the financial returns to higher education quality. In Section 3 we discuss the institutional background as well as the data. Section 4 presents the estimators used in the analysis. The results are available in Section 5 and a discussion on the robustness checks and the implication of the returns is found in Section 6.

II Literature review

The literature on the effect of higher education quality on earnings originated from the U.S.. One of the first studies to account for selection of students between institutions of different quality is Brewer *et al.* (1999). They conclude that even correcting for selection into the type of university attended (identified by net tuition costs), prestigious private institutions provide significantly higher financial returns compared to low cost public institutions, but there are little returns to attending an elite public institution. Black and Smith (2004) confirm that fee differentials are in line with quality differentials, but criticise the parametric approach adopted in the rest of the literature, especially the linearity assumption, and instead recommend using propensity score matching. Their main results pertain to the wage differential between graduates from institutions in the top and bottom quarter of the quality distribution. This large quality gap makes the assumption of selection on observable potentially problematic, and may not represent the typical choice of students who are more likely to arbitrate between institutions of more similar quality. They report premium of 12% for men and 7% for women for attending a better quality institution. When estimating the wage differential between students from the second and third quality quartile

compared to the lowest one, the estimated quality effect was small and statistically insignificant for men and around 12% for women.

However, students' unobservable characteristics may still bias these estimates upwards; for example, if more motivated students attend more prestigious institutions and also, independently of the institution quality, earn higher wages. Dale and Krueger (2002, 2014), use information on all applications to a selection of high quality institutions, linked to Social Security Administration records. To control for selectivity on unobservable, they compare students who applied to the same institutions but went to different colleges. They find no financial return to attending a more selective institution, maybe due to the homogeneity of institution quality in the dataset used (30 highly selective institutions only). They report substantial returns to quality for ethnic minority students, which they reckon, could stem from these students obtaining a network of peers which boosts their career prospect. Hoekstra (2009) uses data on all applicants at a large state university and identifies the quality premium from a regression discontinuity; i.e comparing the earnings at age 28 to 33 of applicants that barely made it, to those of applicants that just failed. The earning premium reaches 20% but this may have little external validity. An alternative strategy to account for unobservable characteristics is to rely on twins who attend different institutions. Behrman et al. (1996) find significant wage differentials between female twins who attended colleges that differ along various measures of quality⁶.

With the exception of Dale and Krueger (2002, 2014) all studies report significant premium to attending a more prestigious institutions, and that OLS estimates are positively biased by selection. It is however unclear how much can be

⁶ Smith (2013) found that attending an institution with a 100 point higher median SAT score, increases the probability of graduating by 5%. This estimate is similar when estimating a between twins model.

extrapolated from US evidence since in other countries the market for higher education tend to be more regulated. Evidence from other countries is scarcer. Papers using administrative registries from Sweden and various identification strategies report positive effect of institutional quality on graduate earnings (Lindahl and Regnér, 2009; de Luna and Lundin, 2014)⁷. However, no returns to the selectivity of the institutions are found in Australia, Finland or Japan (Birch et al, 2009; Suhonen 2012, Nakamuro and Inui, 2012, respectively).

Amongst British evidence, Hussain et al. (2009) exploit various graduate cohorts and a set of quality variables. When combining all measures of quality, an increase of one standard deviation in quality is associated with a wage premium ranging from 2.5% to 5.5%, increasing for the more recent cohorts. They account for selection by including measures of student's ability. Chevalier and Conlon (2003) use propensity score matching to estimate the effect of university quality on the earnings of three cohorts of graduates (1985, 1990 and 1995). Their measures of quality are an indicator of appurtenance to a self-selected pressure group of prestigious universities (Russell Group) and an indicator for when the institution was granted university status⁸. Graduating from the most prestigious institution is associated with a wage premium ranging from 1% to 6% but there is no significant difference between the earnings of graduates at old and new institutions. There is also some evidence that the premium for attending the higher quality institutions increases for more recent cohorts suggesting that as the number of graduates expanded employers may have used institution prestige to differentiate between candidates. Chevalier and Conlon (2003) also estimate the effect of institution quality on wage growth and reports that

⁷ Lindahl and Regnér (2009) use a between-sibling estimates to reduce the bias due to unobserved characteristics and estimate that in Sweden OLS estimates are twice as large as the within family estimates, suggesting a large bias in regressions not correcting for selectivity.

⁸ Eliasson (2007) uses the same dichotomy as a proxy for institution quality in Sweden. Propensity score estimates reveal no effect of institution quality on earnings four to nine years after graduation.

the effect of institution quality on earnings is stable (for the first 10 years after graduation). Broecke (2012) uses administrative data for a cohort of school leavers and their full record of university applications. He matches students who were accepted to the same institution but were one failed to gain the grades to satisfy the conditional offer and reports that a one standard deviation improvement in quality leads to a 7% increase in wages 3 years after graduation.

Finally, Black and Smith (2006) also highlight that the estimate of quality is likely to be biased downwards by measurement error. Most studies have estimated quality by relying on the average SAT scores of attending students, but since quality is likely to be multi-dimensional this is only a proxy for the institutional quality. Instead, Black and Smith (2006) recommend using an array of quality measures and/or estimation techniques that account for measurement error (IV, GMM, Bounds).

III Institutional background and data description

The Longitudinal Destination of Leavers from Higher Education (LDLHE) was conducted in November 2006 amongst a random sample of students who graduated in the summer of 2003. The survey is conducted in two stages. First, the universe of all higher education leavers is sampled six months after graduation (75% response rate). In the second stage, a 55,900 sample is selected to take part in the longitudinal study. The response rate at this second stage is typical of postal survey (44%) and Tipping and Taylor (2007) provide evidence in favour of the representativity of the survey. Survey weights are used throughout the analysis.

The LDLHE is linked to administrative data from the Higher Education Statistical Agency so that additional information on secondary schooling and

university achievements, as well as family background can be added. We select first degree holders, aged 18 to 25 on graduation, non-special entry students and who are currently observed in employment, with valid earning information. Since hours of work are not available in the survey, the labour market outcome of interest is annual earnings. We thus drop part-time workers. We also drop 21 individuals with self-reported earnings above £60,000 (in 2006) three years after graduation so as to reduce measurement error. This leads to a sample of 6,986 observations. (See Table A1 for details on the sample selection).

We supplement the dataset with measures of institution quality. There is no unambiguous measure of institution quality in the UK. All major newspapers provide annual university ranking differing in their methodology. We use information collected from the “Good University Guide”⁹, the longest running provider. Rather than using the ranking we compute a quality measure based on the first principal component along five dimensions of quality for 113 institutions: Research Assessment score, student/staff ratio, academic expenditures per student, mean entry grade and graduate prospect, which we then normalise to a mean of 0 and a standard deviation of 1. The first three variables measure the quantity and quality of the input, the mean entry grade is a measure of the selectivity and popularity of the institution but also captures the quality of peers. This has been widely used as a measure of institution quality. The last measure of quality is the probability that graduates are in further studies or graduate employment six months after leaving university and reflects the employers/universities view on the institution quality. The variables

⁹The Good University Guide is one of the providers of ranking of universities. Rather than using its ranking, we only use the raw variables which can be obtained from: <http://www.thegooduniversityguide.org.uk>. 48 institutions not reporting all the covariates used to compute the quality score are thus excluded. They are mostly small specialised colleges and represent 876 observations.

pertain to data collected for the academic year 2000, when the students applied and the 2001 Research Assessment Exercise. The quality is computed at the level of the institution and not specifically for the subject studied. It can be argued that this is the appropriate level of aggregation to measure higher education quality for labour market related outcomes since employers may have knowledge of institutional quality but not subject specific quality. The first principal component accounts for 70% of the variation in quality. The resulting ranking appears plausible with the top five institutions (in alphabetical order) being: Cambridge, Imperial College London, London School of Economics, Oxford, and University College London. It also compares favourably with other measures of quality that have been used. For example, the mean quality score is significantly different for Russell group institutions (2.42), a group of the most prestigious institutions, the “group 1994”, a second tier of research institution (1.42) and the remaining institutions (-1.14). Indeed, the distribution of quality appears bimodal (Figure 1) and is characterised by a long tail at the high quality end. The lower part of the distribution is mostly formed of the new institutions which were granted university status in 1992.

[Figure 1: Here]

It is unclear what the driver of the quality differences are. UK institutions are charities whose teaching activities are funded by a block grant and tuition income. The grants are a function of the number of students weighted by subject types. In England, tuition fees are capped for home and EU students (currently at £9,000) but are uncapped for non-EU students. As such, teaching funding differs by the mix of subject offered and the share of non-EU students. The latter varies from 0% to 40% and may be a driver of quality differences. The other sources of incomes are research income, which also varies dramatically by institutions – the Russell group institutions

for examples accrue 75% of the research grants income¹⁰; however the extent that this income is used towards teaching is unknown. Endowment and other type of charitable contributions are very small, representing less than 1% of the sector income and are unlikely to generate variations in funding between institutions (UniversitiesUK, 2011).

Figure 2 confirms that in the absence of any additional control there is a positive relationship between the institution quality and the average earnings of graduates. In fact, two clusters can be distinguished. At below median quality, there is little variation in earnings between institutions. For institutions with positive quality score, the relationship between quality and graduate earnings is much stronger. The graph thus highlights that the relationship between institutional quality and graduate earning is convex, rather than linear. OLS estimates, which rely on a linearity assumption between quality and earnings may thus be biased. The differences in wages by quality are substantial. Moving from an institution in the bottom tiers to a top five institution is associated with 50% higher wages.

[Figure 2- Here]

The second relationship that we need to investigate is the amount of sorting between the academic ability of students and the quality of institutions. As stated previously, if the sorting was perfect it would not be possible to identify the effect of institution quality. However, the allocation mechanism of students to institutions in the UK, offers scope for some heterogeneity in the composition of the student body by institution. The admission process to the 134 universities is centralised. From the autumn to the spring preceding their admission to university, high school pupils fill a

¹⁰ www.russellgroup.ac.uk/research - accessed on 25th July 2014.

standard form on-line at the Universities and Colleges Admissions Service (UCAS). Prospective students can state a maximum of six choices (institution, subject) on this application¹¹. Since students have imperfect information on their ability (Chevalier et al. 2007 or Furnham, 2001, for a review), the restricted number of applications means that one strategy to guarantee access to higher education is for some students to apply to institutions spanning the quality range. UCAS then send this form to all chosen institutions/departments which decide whether or not to make an offer to the prospective students. As such, all institutions decide simultaneously, based on exactly the same information. Note that at this point, candidates have not yet sat their end of secondary education exams (usually A-levels). Instead, the decision to offer a place at an institution is based on the teacher's prediction of grades at this high stake exam and previous grades. The offers are typically conditional on applicants reaching a pre-determined score in their, still to come, high stake exams¹². When all institutions have responded, the applicant has to keep only one offer, and choose another one as an insurance choice, in case she does not reach the standards required by her preferred conditional offer.

In August preceding the start of the academic year, the high-stake exam results are revealed. Conditional offers are confirmed or terminated. There is no possibility to trade-up, i.e. candidates who over-performed are locked in their accepted offer, and may thus be observed in institutions of lower quality compared to

¹¹ Students applying to medical, dentistry and veterinary schools, as well as candidates to Cambridge and Oxford universities, who can apply only to one or the other institution but not both, can only state four choices. These applications need to be received earlier on in the cycle, typically in the October of the year preceding their entry to higher education. Since 2008, the number of choice has been further restricted to 5. There is also a slightly different application process for performing arts courses.

¹² Applicants can also apply with other qualification than A-levels. UCAS creates a score to harmonise grades at different type of qualifications. For the remaining of the paper, we will refer to the score at the exams warranting entrance at the university as the entry score.

their peers whose high exam performance was correctly predicted¹³. Pupils who fail to achieve the requirements of their first choice are committed to their insurance choice. For those who also fail the requirement of their insurance choice, a clearing mechanism allocates candidates to institutions that still have places on their courses. About 10% of successful applicants gain access through clearing. These clearing places are allocated on a first come first serve basis – as long as academic credentials are appropriate. A candidate who was accepted in a high quality institution but marginally failed to achieve the required score may at this stage only be able to register in a lower quality institution. Institutions which under-recruited may at this stage lower the recruitment threshold, leading to lower ability students being accepted to high quality institution. This allocation mechanism of students to institutions makes it credible that while selection is based on academic merit, it is imperfect and creates the potential for mismatch between student's ability and institutional quality.

The LDHLE is linked to administrative data so that we observe the same information as universities do when deciding whether or not to make an offer, with the exception of the entry score, for which we observe the realisation rather than the prediction¹⁴. However, since realised scores determine whether the offer is upheld or not, they contain the relevant information to match students to the institution they attend.

Average attainment differs widely between students at colleges of different quality. At the top quartile institutions the average entry score is 23.60 out of 30, 6

¹³ The possibility to continue searching after having an offer confirmed will be offered to applicants in 2014.

¹⁴ The compulsory fields on the UCAS form are name, gender, age, address, country of birth, nationality, financing, disability, ethnicity, occupational background of parental figure if under 21 and previous school attended, which we also obtain when linking the survey to administrative data. The UCAS form also includes previous qualification, references – typically for the pupils' teacher – the student personal statement and their predicted exam results if still currently in high school. This information is not available in our dataset, and instead we have the realised score. We do not know which institutions a candidate applied to, nor what their decision was.

points more than in the third quartile and 14 points more than in below median institutions (Table 1). Figure 3 plots the distribution of entry score by quality quartile. Clearly the allocation of students to institution is not random; and the distribution of test scores shift to the right for every quartile of institutional quality. The top quality institutions recruit almost no students with below average ability, and have a more compressed distribution of ability; this is expected since low performers would not have received an offer if correctly predicted or would have seen their conditional offer elapsed. However, a small fraction of high ability students are found in the lowest quality institution. The sorting is thus asymmetrical which is consistent with the recruitment procedure highlighted above. While common support is universal, the support is thin in the tail of the distribution. These conclusions of asymmetric sorting and thin common support are similar to Black and Smith (2004) for the US.

[Figure 3 – Here]

Students at institutions of different quality differ along other dimensions too. Table 1 reports the means for all variables reported by applicants on their application form. Students at the highest quality institutions have in general more favourable characteristics. For example they are four times less likely not to have a reported entry score, three times more likely to have parents in manager or professional occupations, and 10 times more likely to have been educated at a private school than students at the lowest quality quartile institutions. Living with parents while studying reduces the choice of institutions that a student can apply; students at the top institutions are three times less likely to be living with their parents. The differences in student characteristics between institutions are thus consequent.

The survey has only one measure of income: self-reported yearly income three years after graduation. As highlighted in Figure 1, institution quality is associated

with wage differential and students from the top quartile institutions earn 25% more than their peers at the lower end of the quality distribution. The wage gap appears especially large between 3rd and 4th quality quartiles. Moving from quartile 1 to 2 and to 3 is associated with an average wage increase of £1,000 but moving from 3 to 4 the gap is more than three times as large. Quality effects may thus be heterogenous.

IV Empirical strategies

We conduct the analysis using several empirical strategies. The first is to estimate the effect of institution quality on log earnings ($\ln Y$) by Ordinary Least Squares. Typically, the estimated model would be of the form:

$$\ln Y = \beta_0 + \beta_1 Q + \beta_2 X_1 + \beta_3 X_2 + \varepsilon \quad (1)$$

where Q is a measure of university quality, X_1 a set of individual characteristics on graduation, including a measure of the student pre-enrolment ability, and X_2 a set of current characteristics. X_2 can be considered endogenous as the university quality may affect various dimensions of the labour market attainment of graduates. In the application, X_2 contains only indicator of the regional labour market conditions (126 post code for work location). Note that standard errors are clustered at the institution level to account for possible correlations between observations.

The parameter of interest is β_1 which represents the increase in earnings due to an increase in university quality. As stated previously β_1 may be biased, even if (1) includes all confounding factors correlated with both quality and earnings, if the relationship between quality and earnings is non-linear and if there is a lack of common support (in which case the identification is purely due to the imposed functional form). The first limitation can be eliminated by measuring quality in a non-linear form: a set of dummies, or a polynomial function for example.

(1) can also be estimated by quantile regression in order to test for heterogeneity in the effect of institutional quality on earnings which have been highlighted in the descriptive statistics. So that for each quantile p of the wage distribution we can, following Koenker and Basset (1978), define the conditional quantile of the distribution as:

$$Q^{(p)}(\ln Y / Q, X_1, X_2) = \beta_0^{(p)} + \beta_1^{(p)} Q + \beta_2^{(p)} X_1 + \beta_3^{(p)} X_2 + Q^{(p)}(\varepsilon) \quad (2)$$

and impose the restriction that $Q^{(p)}(\varepsilon) = 0$ to estimate the quartile specific parameters.

To test for common support, we also estimate model (1) by propensity score matching (PSM), where, quality is measured as a binary variable. The parameter of interest is then the mean differences in wages between graduates who attended a high quality institution compared to those who did not but who, based on their observable characteristics, could have done so, this can thus be considered an Average Treatment on the Treated (ATT). If we define Y_1 the wage if attended a high quality institution and Y_0 the wage if attended another institution, and D is an indicator of having graduated from a high quality institution then the parameter of interest is simply:

$$ATT = E(Y_1 - Y_0 / D = 1) \quad (3)$$

ATT can be estimated if $E(Y_0 / D = 1)$, which is never observed, can be approximated. If the allocation of students to universities is not random, then $E(Y_0 / D = 1) \neq E(Y_0 / D = 0)$. However, if it can be argued that the selection is based on observable characteristics (X) then the earnings of non-treated individuals can be used as counterfactual to approximate the unobserved earning of treated individuals if they had not been treated, i.e. conditional on this set of variables X an observation can be considered randomly allocated to either the treatment (high quality institution) or

the control group ($E(Y_0 / X, D=1) = E(Y_0 / X, D=0)$). Formally, this is expressed in the Conditional Independence Assumption: $(Y_0 \perp D / X)$. Since we observe almost the same information as universities do when they make an offer to a student the assumption of selection on observable is plausible. The only difference in the set of observables is that universities make their offer using the results of past exams and the predicted grades at A-levels while we observe the realised A-levels grades. We do not have access to the predicted grades, nor the breakdown of the entry score by subjects. However, since the realised grade, not the predicted one, guarantees the access to higher education it is in fact the most relevant information to base our matching.

Both OLS and matching produce unbiased estimates as long as the selection into institution of high quality is due to observed variables, but matching also highlights potential bias due to lack of common support. Formally, the common support can be expressed as: $0 < \Pr(D=1 / X) < 1$; individuals with the same characteristics have a positive probability to be treated or non-treated. The probability of treatment is the propensity score which is used to match observations, rather than X , so as to reduce the dimensionality of the matching problem. Rosenbaum and Rubin (1983) show that it is equivalent to match on all the components of X or on $\Pr(D=1/X)$.

For each treated individual i , the counterfactual outcome associated with no treatment is a weighted average of the outcomes from control observations, where the weights are a function of the distance between the treated and control observations' propensity scores (using Epanechnikov kernel). If no match is found within a bandwidth, the treated observation cannot be matched which highlights the lack of common support discussed above. Black and Smith (2004) estimate the effect of moving from the bottom quartile of the quality distribution to the first quartile. Only a

fraction of students will consider switching from a bottom to a top quality institution, and the estimate may be of limited interest. Moreover, by considering institutions that are so different, the common support always become extremely thin. Instead, we estimate three effects: moving from the first to the second quality quartile, second to third and third to fourth.

Finally, since quality is (almost) a continuous variable we extend the propensity score matching estimator to the case of continuous treatment (Hirano and Inbens,2004). The remaining part of this section draws heavily on this article. The intuition behind Continuous Propensity Score Matching (CPSM) is rather similar to the dichotomous case. A Generalised Propensity Score (GPS) is calculated which is the density of the treatment (T) conditional on a set of covariates X: $R=r(T,X)$, the GPS has similar properties to the propensity score in the dichotomous case, mainly it satisfies CIA, so that: $X \perp 1\{T = t\} | r(t, X)$.

A two-step procedure is then used to remove any bias due to differences in X. First, we estimate the conditional expectation of the outcomes as a function of the treatment (T) and the GPS: $\beta(t, r) = E[Y/T = t, R = r]$. This regression function does not have a causal interpretation and is only used to calculate the dose response function (μ) in the second step. $\mu(t) = E[\beta(t, r(t, X))]$. Computationally, we estimate the GPS by OLS ($T/X \sim N(\beta_0 + \beta_1'X, \sigma^2)$) and calculate:

$$\hat{R} = \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} \exp\left(-\frac{1}{2\hat{\sigma}^2} (T - \hat{\beta}_0 - \hat{\beta}_1'X)^2\right) \quad (4)$$

In the second stage we estimate, again by OLS, the outcome as a function of a flexible function of the treatment and the GPS. Following Kluve et al. (2012) we estimate:

$$E[Y | T, \hat{R}] = \alpha_0 + \alpha_1 T + \alpha_2 T^2 + \alpha_3 T^3 + \alpha_4 \hat{R} + \alpha_5 \hat{R}^2 + \alpha_6 \hat{R}^3 + \alpha_7 \hat{R} * T + \alpha_8 \hat{R}^2 * T + \alpha_9 T^2 * \hat{R} \quad (5)$$

(5) is then used to estimate for a given level of treatment t , the average potential outcome. The dose response function is found by computing this last stage for all level of the treatment considered. Hirano and Imbens (2004) recommend bootstrapping to obtain standard errors of the estimates. Rather than the dose response function, we report its derivative: in our case the interpretation is the marginal return to university quality. We also compute the expected (log) earnings at all values of the treatment.

As in the case of dichotomous treatment, the GPS is only valid if the matching leads to a balancing of the characteristics of the treated and untreated group; i.e. akin to an-post random allocation of treatment. To assess the balancing we follow two strategies. First, as in Hirano and Imbens (2004), we divide the institution quality variable into three terciles and test, for each variable, whether the GPS adjusted mean differs in one tercile compare to the other two. This is equivalent to testing that the conditional mean and the treatment indicator are independent (CIA) where $r(t, X)$ is evaluated at the median value of the treatment within the tercile (t^*). We follow Hirano and Imbens (2004) and test this hypothesis by blocking. For each tercile, we define five blocks defined by the quintile of $r(t^*, X)$. For each block we calculate the mean difference in X for observations ($T = t$) and ($T \neq t$), and combine these five mean differences, weighted by the number of observations in each block, to calculate a t-statistics of the statistical differences in the mean of X between treated and untreated observations in that tercile.

Imai et al. (2008) suggest that such a test may have low power as the decrease in the t-statistics may be due to an increase in the variance rather than a reduction in

the difference in means (Balance Test Fallacy). Thus, we also conduct a second balancing test. Imai and van Dyk (2004) recommend to test the balancing of the covariates by regressing each covariate on the treatment variable and the predicted treatment $E[T|X]$ and testing for the significance of the treatment. An alternative to this test is to regress each covariate on the treatment and the GPS, at different values of the GPS (Kluve et al. 2012).

V Results

As discussed above, the main criteria used by universities to accept applicants are their predicted score at a high stake exam but we use the realised score instead to control for students' ability¹⁵. The decision to apply to universities of different quality may also be related to individual characteristics (sex, age, ethnicity, disability status) and other background characteristics (private school, parental occupational class, whether expect to live at home while studying and tuition fee status) which we include in the earning regression or in the matching process. All these variables are available to universities when assessing a student's application. We additionally control for region of origin and region of the institution attended.

V-A Discrete measure of quality

First, we recode the measure of university quality into a discrete measure to facilitate comparisons with previous research. Quality is then measured in quartiles of the continuous quality measure with the fourth quartile representing the highest quality. Table 2 reports estimates of (1) estimated by OLS for different specifications

¹⁵ Gibbons and Chevalier (2007) find small differences between teacher assessments and realised scores at age 16 especially for pupils at the extremes of the ability distribution but these differences have no impact on subsequent pupil outcomes. Hayward et al. (2005) report "reasonable accuracy" (+/- one grade) at A-levels and no impact of the error on higher education participation.

of the covariates. In the first column, the only additional control is a set of postcode dummies to capture the characteristics of the local labour market. There are some marked institution quality effects and graduates from the top quality institutions earn 20% more than those from the lowest quality institutions. This premium is only 8.4% for graduates from the 3rd quality quartile and no significant premium is observed for graduates from the 2nd quality quartile. However, these estimates are biased upwards since the ability of students is correlated both with institutional quality and earnings. Thus, the second column adds a cubic function of the entry score to control for ability. The estimates on the effect of quality are reduced by at least a third, confirming that there is a large amount of sorting of students by ability. Note that for students attending a 2nd quartile institution, controlling for attainment does not alter the returns; i.e. the intake at these institutions are similar. The next column reports estimates that controls for various dimensions of the individual characteristics and socio-economic background, including region of origin and region of institution; this reduces the quality effect further, so that the premium to attending a top institution is halved compared to the original specification. The fourth column adds controls for subject of studies and degree grades since institutions of different quality may differ along these dimensions as well. Indeed there are marked differences in the distribution of subjects by quality of institution (Table A2). Medicine, all sciences and engineering degrees, law, Classics, Languages, history, Economics, are more likely to be taught at higher quality institutions. While Business, Communication, Creative Arts, IT, being more popular at lower quality institutions. Controlling for subjects (and final grades) improves the precision of the estimates and reduces the gap further, so that they now range from 2.5% to 8.0%. Despite the rich set of controls significant quality differentials remain for institution above the median quality, and especially for

those in the top quality quartile, which suggests that differences in earnings cannot be fully explained by selection.

[Table 2 Here]

We now estimate the institution quality effects by propensity score matching for three different treatments: attending an institution in the 4th quality quartile rather than one in the third; a third quality quartile institution rather than a second; and a second quality quartile rather than a first. These treatments are likely to be similar to students' alternatives when selecting universities. Propensity scores are estimated separately for each treatment. In each case, the controls are a cubic function of entry score, an indicator for having no reported score, gender and disability, dummies for age and ethnicity, proxies for financial constraints, subject dummies, and dummies for region of origin¹⁶. The matched samples are balanced for the treated and control groups on all characteristics. The common support is almost universal, between 94% and 99% of the treated are matched to some control observations but, the support tends to be thin for high value of the propensity score - see Figure 4 for an example when the treatment is attending a 4th quartile university rather than a third quartile university. Figures for the other treatment are similar.

[Figure 4 Here]

The PSM estimates are reported in Table 3. The results confirm that returns to quality are only observed for above median institutions. Wages are 5% higher for graduates from the 4th quartile over those from the 3rd quartile. Those graduates also enjoy a premium of 4% over those who attended an institution from the 2nd quartile. There is no difference in earnings between graduating from the bottom two quality

¹⁶ It was not possible to control for region of institution since for some regions, there is only one institution in a given quality quartile.

quartiles. The OLS and matching estimates lead to the similar conclusions; the effect of institution quality on earnings is non-linear, and only exists for above the median institutions. Note that the returns to quality are substantially larger when estimated by propensity score matching. For example, the returns to attending a 4th quartile rather than a 3rd quartile institution is 3.4 percent when estimating by OLS (0.08-0.046) but 5.3% when estimated with propensity score matching.

V-B Continuous measure of quality

Rather than relying on discrete measures of quality the analysis is now conducted using a continuous measure. The lower panel of Table 2 reports the estimates on a quadratic function of quality for various specifications¹⁷. In the most parsimonious model, moving from one standard deviation from the mean quality increases earnings by 7.7%. However, as in the discrete case, adding further covariates considerably reduces this premium and in the most complete model this premium has been more than halved and reaches 3.4%. This estimate is doubled for an institution with a normalised quality score of 2, highlighting the non-linearity of the returns. A specification using a log quality measure was also tested, the estimated elasticity of quality then ranges from 0.12, for the most parsimonious model, to 0.05, for the full control one. Later we investigate the non-linearity of this relationship without imposing a functional form.

Relying on a central tendency measure may be misleading. We thus additionally report estimates from quantile regressions. Rather than imposing that the effect of university quality is homogenous at all level of income, quantile regressions allows it to differ for different quantile of the income distribution. One may for example,

¹⁷ A cubic function of institution quality was also estimated but neither the quadratic nor the cubic terms were ever statistically significant.

expect that individuals with higher earning potential may gain more from attending a higher quality institution. To simplify the interpretation of these estimates those estimates are based on a linear function of quality.

For each decile the estimated effects are reported graphically in Figure 5. For the first six income deciles, the estimated effect of university quality lies at the bottom of the OLS confidence interval and average 0.046. The estimated returns to quality then jumps almost 1 percentage points to the OLS estimate for the seventh decile, and thereafter becomes a positive function of the income decile. However, only the estimate for the 9th decile (0.078) is significantly different from the OLS estimate. Thus, there is a strong heterogeneity in the returns to institution quality but only for individuals at the top of the earnings distribution.

[Figure 5: here]

The previous estimates using a continuous measure of university quality have assumed a random allocation of students to universities. We keep assuming that the allocation is based on observable characteristics but use the Continuous Treatment Matching method proposed by Hirano and Imbens (2004) to estimate the institution quality effect on the earnings of recent graduates, so that no functional form on the relationship is assumed. We first estimate the GPS by regressing a log transformation of the quality score on all the covariates used previously and computing GPS value using the normal distribution (4)¹⁸.

The balancing properties of the estimated GPS are assessed in Table 4. Following Hirano and Imbens (2004), we first split the sample into three terciles with respect to the quality of the institution attended. The first three columns of Table 4, report for each covariate a t-test of the mean difference in one tercile (g) compare to the other

¹⁸ We report results when a log transformation of the quality score is used since the log transformation is closer to a normal distribution which we use to compute the GPS. Estimates based on the untransformed quality score were also computed and do not substantially differ from those presented.

two groups. Clearly, there are marked differences between graduates from institutions at different level of the quality distribution, highlighting that selection of students take place. We then compute the GPS at the median value of the tercile and split the group into 5 blocks. Within each block, we compute the t-statistics of the difference in means between treated (went to an institution of quality g) and non-treated observations. Columns 4 to 7 report the weighted average t-statistics for each block. After conditioning on GPS, most mean differences between groups are dramatically reduced and insignificant. There remains a small significant difference in the highest A-level score category, parental background, school type and for medical graduates. We fare less well in balancing on regions of origin and region of origin with 9 tests being positive. Overall, we can still be pretty satisfied of the balancing, out of the 297 tests conducted only 15 are significant, which is exactly the number of false positive expected with a 95% confidence interval.

[Table 4]

Imai et al. (2008) suggest that such a blocking test may be subject to balance test fallacy. As an alternative balancing test, we also conduct the Imai and van Dyk (2004) test. In columns 7, we report for each covariate, the t-tests on the log treatment coefficient in an un-conditional regression. Those are large, indicating that covariates and quality are correlated. In column 8, we report the same t-test when we condition on the predicted value of the treatment. Again we observe a large reduction in the value of the t-test but some remain above the critical value for significance. These concern the top of the ability distribution, school type, some subjects and the regional dummies. Kluve et al. (2012) recommend to run a similar test but conditioning on GPS values at the 25th, 50th and 75th percentile of the quality distribution instead instead. This test broadly confirms the Imai and van Dyk and is not reported here. We

have conducted these three tests for various specifications, including interactions of the various controls, but the conclusions are always the same. The blocking-test suggests that the after conditioning the samples are balanced while the other two tests highlights some balancing issues. Eyeballing the means and standard errors in the blocking test, it does not appear that the reduction in the t-tests when blocking is due to an inflation of the standard error. Since the blocking test assess the balancing over a finer range of the common support, we are satisfied that after conditioning on GPS the sample is balanced.

To test for common support, we follow the procedure highlighted in Kluve et al. (2012). We split the sample into three tercile as previously. We then evaluate the GPS at the median quality value of group g for all observations. We replicate this procedure three times, once for each tercile. Figure 6 reports the distribution of GPS estimated at the median value of group g , for group g versus not group g . For each group, the GPS distributions overlap indicating common support. Note that the common support is thinner for the lowest and highest quality groups.

[Figure 6]

Figure 7A reports the marginal effects of (log) institution quality on earnings throughout the distribution of treatment. The quality effect is convex and positive but for a large section in the lower part of the quality distribution, not actually significantly different from 0. For institutions above the median quality, there are some positive returns to quality. Note that the return is increasing with quality, so that the highest quality institutions generate the highest return to quality. This may be better appreciated in Figure 7B which reports the expected log earnings over the quality distribution. The expected wage differentials are only significantly larger for

attending an institution in the top quartile of the quality distribution. The predicted wage differential between a graduate from the median 4th quality quartile and one from the 3rd quartile is 7%.

[Figure 7]

In Table 5, we provide evidence that these results are robust. First, we eliminate institutions and subjects for which the selection process includes an interview (Oxford, Cambridge) or an additional test (Medical schools, Art schools) as this makes the assumption of selection on observable characteristics less palatable. This selection eliminates both high quality institutions from which there is a high premium to quality, and arts graduates which typically have low earnings. As such, the quality effect remains almost identical to those obtained on the full sample. Second, we assess whether there are gender differences in the returns to institution quality. The gender wage gap reaches £2,000 or nearly 10%. However, there are no significant differences in the returns to quality by gender. In fact, the CTM predicted wages suggest that the gender wage gap may open up for high quality institution graduate. The gender wage gap in predicted wage is around 6% for graduates from the bottom three quartiles but reaches 7.5% for graduates from top quartile institution.

Finally, the last two columns assess the sensitivity of our results to alternative measures of quality. First, we eliminate graduate prospects as this is measuring a post-university outcome. The second alternative is to additionally drop the research quality score, as this may not be directly related to teaching quality. The estimates tend to be somehow smaller, confirming that simpler measure of quality tend to reduce the returns to quality (Black and Smith, 2004). Overall, we are satisfied that our estimates are robust. Additionally, we have experimented with different transformation of the quality measure but the general conclusions remain the same.

[Table 5]

VI Discussion

Using a wide array of estimators, we consistently find that there is a large amount of sorting to institution, the estimated quality effects are halved when including students' characteristics (mostly ability). Secondly, institutions quality is associated with positive financial returns for graduates. While there are some differences on the size and the distribution of the effect, the overwhelming conclusion is that the quality effect is non-linear and accrues mostly to graduates from the highest quality institutions or with the best earnings credentials. The estimated returns to graduating from a top quartile institution as oppose to a third quartile one, range from 3.5% to 5.5% when estimated by OLS or PSM. Our favoured estimates, GPS, are even larger reaching 7% for moving from the median quality of the third quartile to the median value of the fourth quartile.

Graduates from a 3rd quartile institution earn on average £22,785 per annum three years after graduation. We can compute the life time premium that would have been associated with these students graduating from a 4th quartile quality institution instead. On average the difference in quality score between these institutions is just short of 1 standard deviation. Our favoured estimated effect in this quality range is 7% (CTM). Assuming a growth rate of 2% over the life time, a 40-years long participation to the labour force, a discount rate of 3.5% and that the returns to quality are constant over the life-time, the differences in the present value of the life time earnings at 18 between a would-be graduate at a third quartile institution and one at a fourth quartile institution reaches £40,000. Since tuition fees are (almost) identical between all institutions – at the cap value stated by the government – there is no potential arbitrage for students between a high quality institution and a cheaper one,

so £40,000 is a net gain from attending a top quality institution. The current admission system is based on predicted grades and trap applicants in their initial choice. As such an applicant whose grades were wrongly predicted and could not get access to a top quality institution suffers from a large financial penalty over her life time.

VII Conclusion

We show that currently in the UK there is a wage premium to university quality. This premium is non-linear. There is no significant quality effect for institutions in the bottom half of the quality distribution but graduates from the most prestigious institutions earn 7% more than graduates from institutions in the 3rd quality quartile. These estimates are robust to an array of methods, all relying on selection on observables. Compare to other UK estimates, such as Broecke (2012) which accounts for selection on unobservable, our results are much lower, maybe because his sample is composed for higher ability individuals than the one used for this analysis. While smaller, the returns to quality still lead to some important differences in life time earnings. Policy makers may thus worry that the current system of fixed price between all institutions can be considered unfair towards students attending the lower quality institutions. The allocation system could also be revised to either be based on the final secondary education exam, rather than its prediction, or by allowing successful applicants to trade up if their exam results were better than expected. A remaining question is whether this premium stems from an increase in human capital, peers effects or signalling.

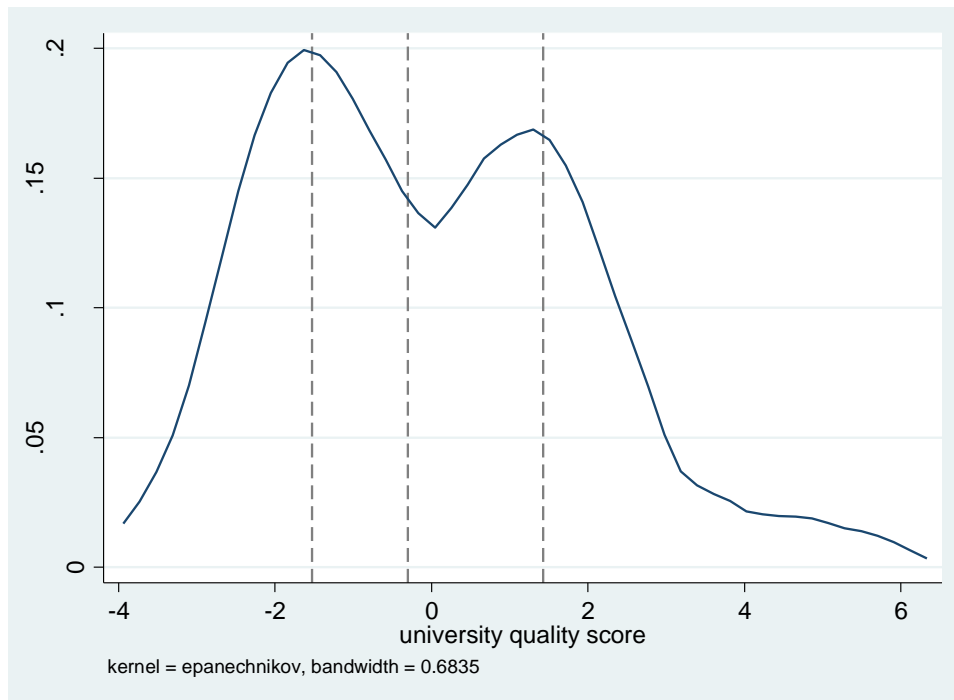
References:

- Behrman J., M. Rosenzweig and P. Taubman (1996) "College Choice and Wages: Estimates Using Data on Female Twins, *Review of Economics and Statistics*, 78, 672-685
- Birch, E. R., Li, I., & Miller, P. W. (2009). The influences of institution attended and field of study on graduates' starting salaries. *Australian Economic Review*, 42(1), 42-63
- Black D. and J. Smith (2004) How robust is the evidence on the effects of college quality? Evidence from matching, *Journal of Econometrics*, 121, 99-124
- Black D. and J. Smith (2006) Estimating the Returns to College Quality with Multiple Proxies for Quality, *Journal of Labor Economics*, 24, 701-729
- Bowman W. and S. Mehay (2002) College Quality and Employee Job Performance: Evidence from Naval Officers, *Industrial and Labor Relations Review*, 55, 700-14
- Brewer D., E. Eide and R. Ehrenberg, 1999, Does it pay to attend an elite private college?, *Journal of Human Resources*, 33, 104-123
- Broecke, S. (2012). University selectivity and earnings: Evidence from UK data on applications and admissions to university. *Economics of Education Review*, 31(3), 96-107.
- Chevalier A. and G. Conlon (2003) Does it pay to attend a prestigious university? *London School of Economics*, CEE, DP 33
- Chevalier A., S. Gibbon, A. Thorpe, M. Snell and S. Hoskins (2007) Students' academic self-perception, *IZA*, DP 3031
- Dale S. and A. Krueger (2002) Estimating the pay off to attending a more selective college: An application of selection on observables and unobservables, *Quarterly Journal of Economics*, 1491-1527
- Dale, S. B., & Krueger, A. B. (2014). Estimating the Effects of College Characteristics over the Career Using Administrative Earnings Data. *Journal of Human Resources*, 49(2), 323-358.
- Economist (2014) "Is College Worth It?", *The Economist*, 05 April 2014
- Eliasson K. (2007) How Robust is the Evidence on the Returns to College Choice? Results Using Swedish Administrative Data, *Umea University*, Mimeo
- Furnham A. (2001) Self-estimates of intelligence: culture and gender difference in self and other estimates of both general (g) and multiple intelligences, *Personality and Individual Differences*, 31, 1381-1405
- Gibbons S. and A. Chevalier (2008) Assessment and Age 16+ participation. *Research Papers in Education*, 23 (2) 113-123.

- Hayward G., S. Sturdy and S. James (2005) Estimating the reliability of Predicted Grades, UCAS, UCAS report
- Hershbein, B. J. (2013). Worker signals among new college graduates: The role of selectivity and GPA. Upjohn Institute, DP 13-190
- Hoekstra, M. (2009). The effect of attending the flagship state university on earnings: A discontinuity-based approach. *The Review of Economics and Statistics*, 91(4), 717-724
- Hussain I., S. McNally and S. Telhaj (2008) University Quality and Graduate Wages in the UK?, *London School of Economics*, CEP, mimeo
- Hoxby, C. M. (2009). The Changing Selectivity of American Colleges. *Journal of Economic Perspectives*, 23(4), 95-118.
- Estelle James, Nabeel Alsalam, Joseph C. Conaty and Duc-Le To (1989) College Quality and Future Earnings: Where Should You Send Your Child to College? *American Economic Review*, Vol. 79, No. 2 pp.247-252
- Imai, K., G. King and E. Stuart (2008). Misunderstandings between experimentalists and observationalists about causal inference. *Journal of the Royal Statistical Society*, series A, 171, 481–502
- Imai, K. and van Dyk, D. A. (2004) Causal inference with general treatment regimes: generalizing the propensity score. *Journal of the American Statistical Association.*, 99, 854–866.
- Jawaguchi D. and W. Ma (2008) The causal effect of graduating from a top university on promotion: Evidence from the University of Tokyo's 1969 admission freeze, *Economics of Education Review*, 27, 184-196
- Kluve J., H. Schneider, A. Uhlendorff, Z. Zhao (2012) Evaluating Continuous Training Programs Using the Generalized Propensity Score, *Journal of the Royal Statistical Society*, Series A, 175, 587-617
- Koenker R. and G. Bassett (1978) Regression Quantiles, *Econometrica*, 46, 33-50
- Lindahl L. and H. Regnér (2005) College Choice and Subsequent Earnings: Results Using Swedish Sibling Data, *Scandinavian Journal of Economics*, 107, 437-457
- Luna, Xavier de, and Mathias Lundin. "Sensitivity analysis of the unconfoundedness assumption with an application to an evaluation of college choice effects on earnings." *Journal of Applied Statistics* ahead-of-print (2014): 1-18.
- Long M. (2008) College quality and early adult outcomes, *Economics of Education Review*, 27, 588-602
- Nakamuro, M., & Inui, T. (2013). *The Returns to College Quality in Japan: Does Your College Choice Affect Your Earnings?*. ESRI Discussion Paper Series.

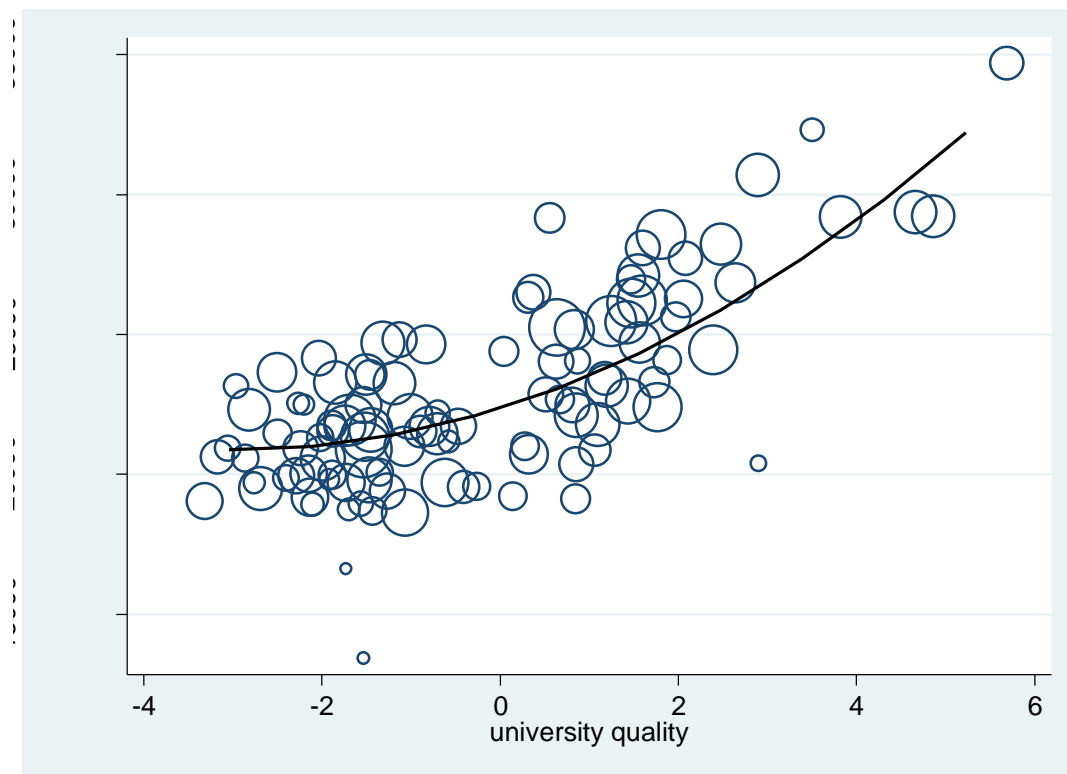
- Rosenbaum P. and D. Rubin (1983) The central role of the propensity score in observational studies for causal effects, *Biometrika*, 70, 41-55
- Robst J. (1995) College Quality and Overeducation, *Economics of Education Review*, 14, 221-228
- Saavedra, J. E. (2008). The returns to college quality: A regression discontinuity analysis., Unpublished Harvard Mimeo
- Smith, J. (2013). Ova and out: Using twins to estimate the educational returns to attending a selective college. *Economics of Education Review*, 36, 166-180.
- Suhonen, T. University Choice and Subsequent Earnings in Finland: Estimates for Groups of Majors. Unpublished Mimeo, University of Jyväskylä

Figure 1: Distribution of Institution Quality:



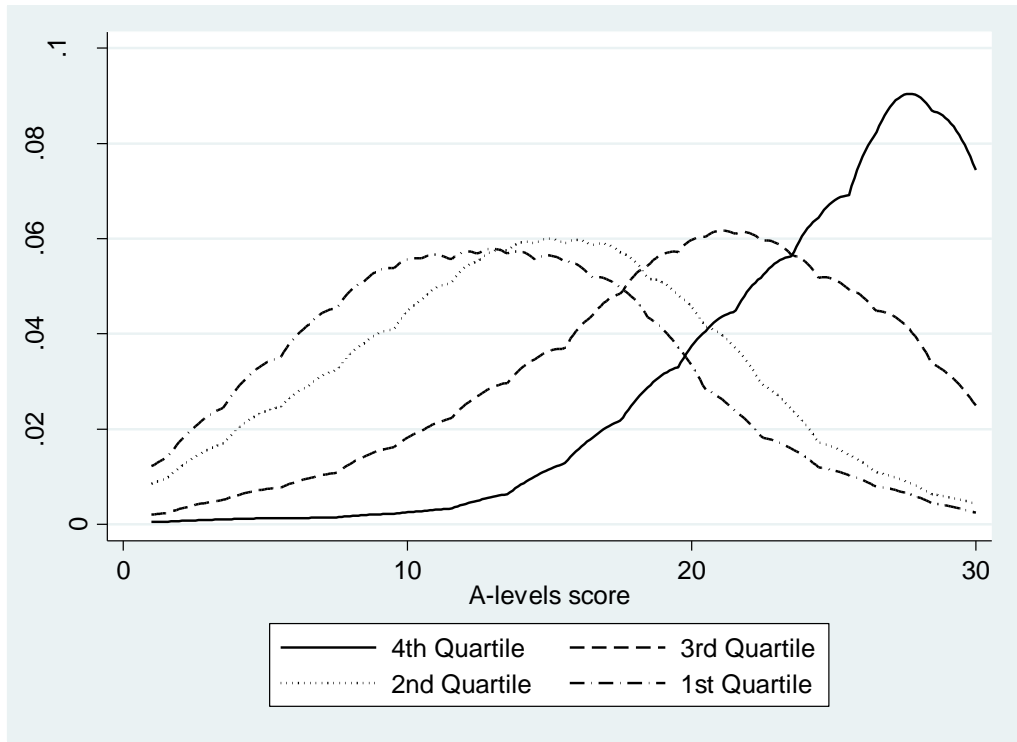
Note: Quality is measured as the principal component of research assessment, student staff ratio, academic expenditure, mean entry grade and graduate prospect. The dash line reflects the thresholds to the 1st, 2nd and 3rd quartile of the quality distribution.

Figure 2: University Quality and Mean Institutional Wage for Graduates



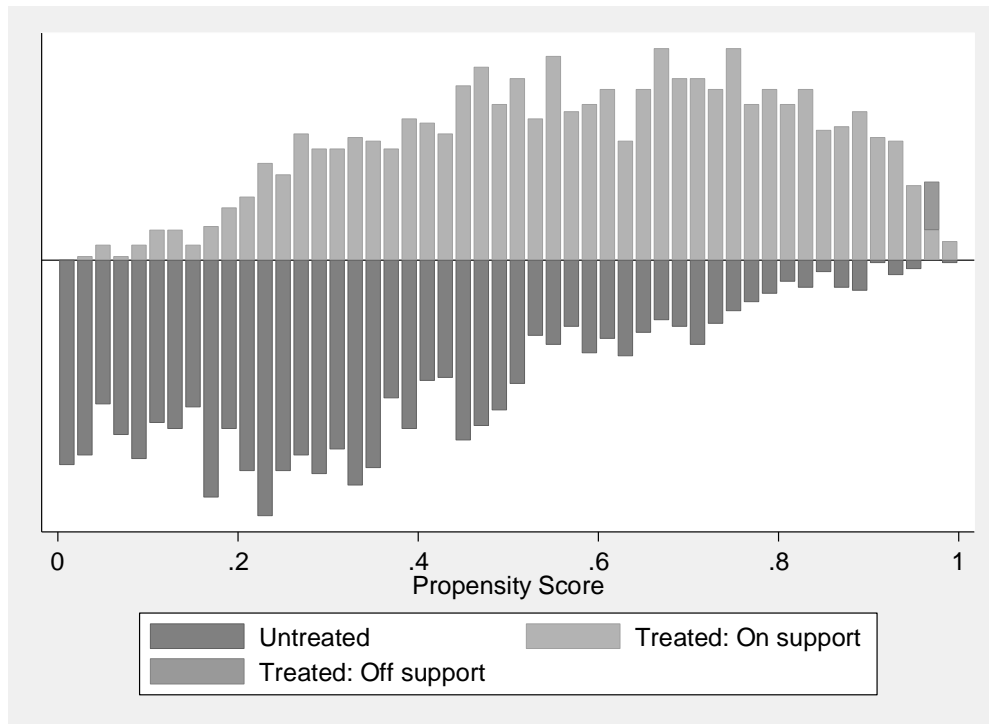
Note: The circle size represents the number of students at the institution. Mean wages are self-reported annual wages for full-time workers earning less than £60,000 three years after graduation. The line is based on a quadratic fit of university quality on earnings.

Figure 3: Distribution of Entry Exam Score by University Quality Quartile



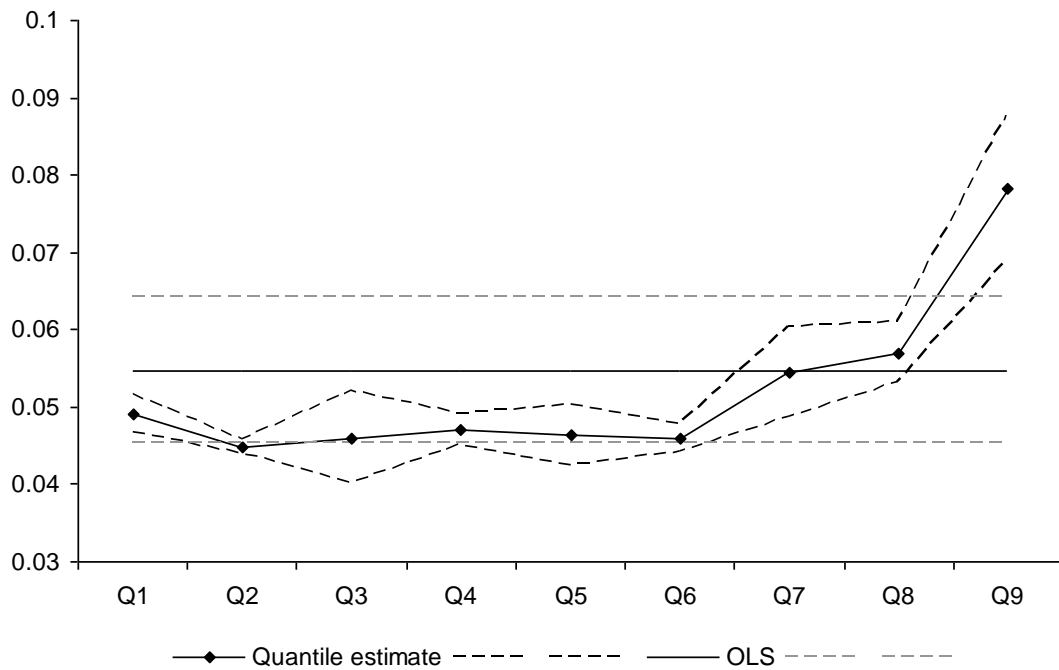
Note: Distribution of entry exam score by quartile of institution quality.

Figure 4: Balancing Propensity Score - Treatment (top quality quartile) and Control (3rd Quality Quartile).



The propensity score includes an indicator of A-level score missing, a cubic in A-level score, gender, parental occupation categories, subject of degree, tuition fees status, accommodation status, disability, type of school attended, ethnicity and a set of dummies for age.

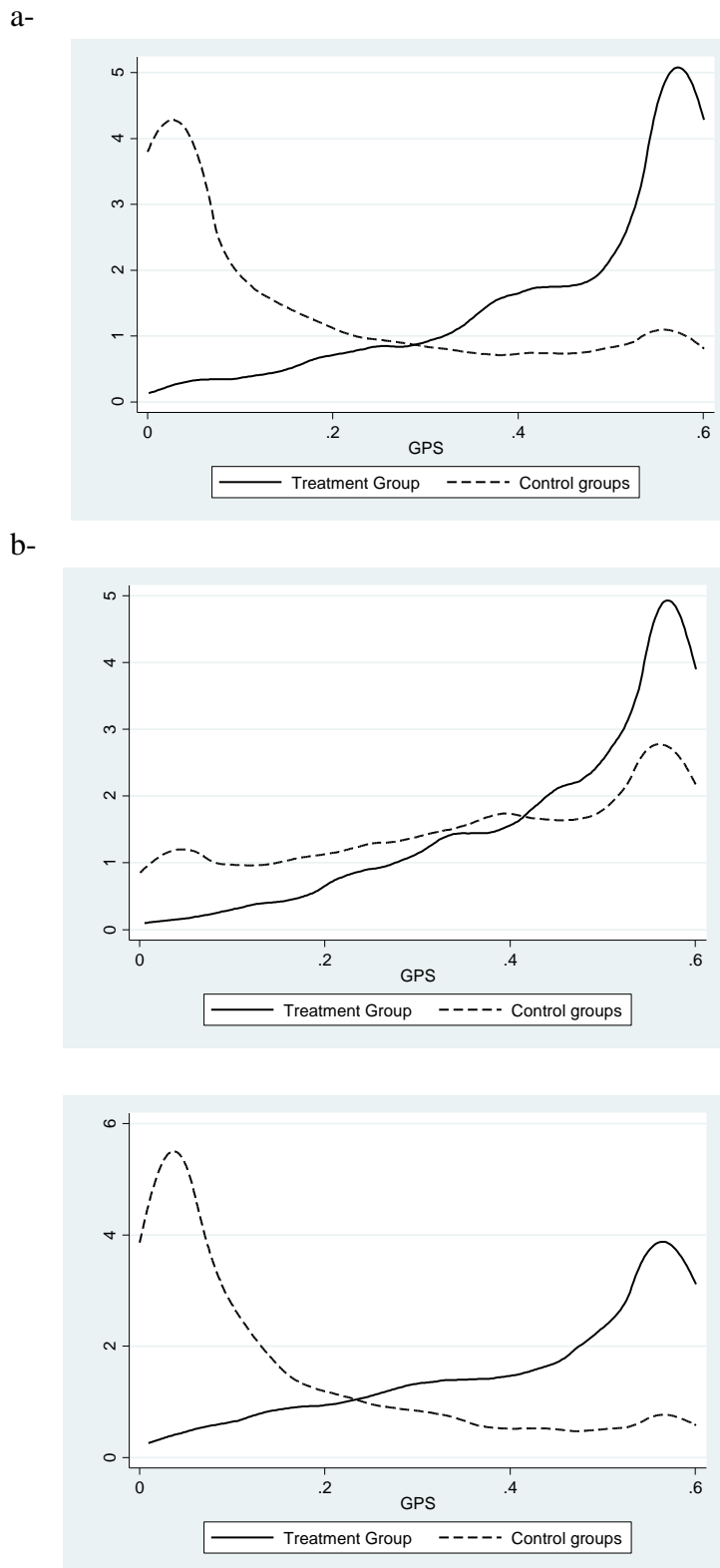
Figure 5: Estimated Returns to Institution Quality: Quantile regressions



Note: Darker dash lines indicate the 5% confidence interval for quantile estimates, grey dash lines indicate the 5% confidence interval for OLS estimate.

Estimated model has the same specification than the most extensive models presented in Table 2, but where the categorical measures of institution quality have been replaced by a linear function of the normalised quality score. Regional dummies have been omitted as small number of observations in some categories prevented the model converging.

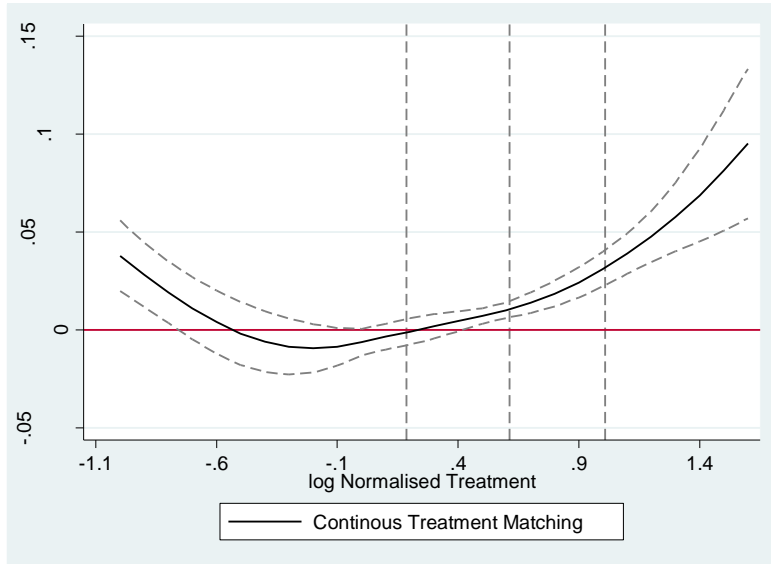
Figure 6: Distribution of GPS – Evaluated at the Median Value of the Treatment for each Quality Treatment Tercile



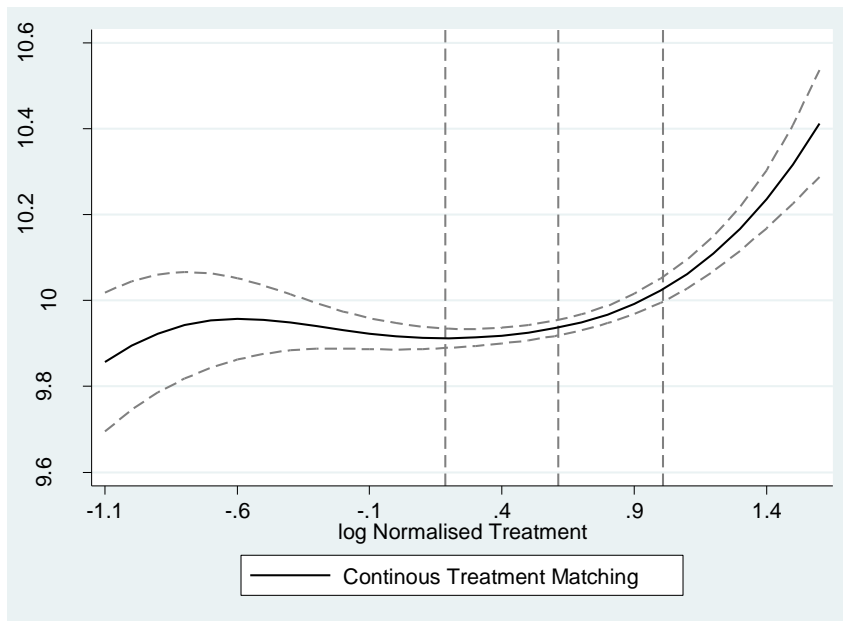
Note: Distribution of GPS estimated at the median university quality for the lowest GPS tercile (a) , medium tercile (b) and upper tercile (c)

**Figure 7: Estimated Marginal Effects of Institution Quality on Earning.
Continuous Propensity Score Matching**

A] Dose Response to Log Quality



B] Predicted Wage to Log Quality



Note: Estimates of the marginal effect of institution quality on earnings from the Generalised propensity score matching. Dash lines represents the 95% confidence interval, obtained from 200 replications bootstrap.

Table 1: Summary Statistics by University Quality Quartile

Variable	Quartile 1	Quartile 2	Quartile 3	Quartile 4
Male	0.42	0.43	0.47	0.45
A-levels score	13.41 (6.21)	15.65 (6.03)	21.49 (5.84)	25.61 (5.03)
A-levels score missing	0.11	0.10	0.02	0.03
Parental occupation				
Manager	0.11	0.17	0.22	0.26
Professional	0.09	0.14	0.22	0.34
Associate professional	0.07	0.07	0.09	0.08
Other	0.28	0.26	0.28	0.23
Not reported/no occ.	0.45	0.35	0.19	0.09
Accommodation				
University	0.13	0.10	0.13	0.20
Parents	0.27	0.24	0.14	0.08
Own	0.31	0.44	0.45	0.62
Other	0.13	0.12	0.11	0.09
Unknown	0.17	0.09	0.17	0.02
Disable	0.04	0.06	0.04	0.03
White ^A	0.78	0.84	0.88	0.86
Age on graduation	20.31 (1.54)	20.35 (1.53)	19.79 (1.11)	20.00 (1.15)
Labour market outcomes				
Ln salary	20,683 (6,113)	21,005 (6,183)	22,785 (7,403)	26,655 (9,112)
Observations	1658	2001	1762	1551

Note: Cells report means of variables and standard error in parentheses when the variable is continuous. The sample is weighted to account for survey structure. Only individuals currently working full-time and reporting earning less than £60,000 per annum are selected in the sample.

^A the analysis is conducted at a more disaggregated level and ethnicity is broken down between Black, Indian, Pakistani/Bangladeshi, other Asian, mixed, other ethnicity.

Table 2: OLS Financial Returns to Quality

Discrete measure	(1)	(2)	(3)	(4)
Quality 2	0.025 (0.020)	0.026 (0.019)	0.028* (0.017)	0.025* (0.014)
Quality 3	0.084*** (0.021)	0.065*** (0.021)	0.053*** (0.019)	0.046*** (0.015)
Quality 4	0.205*** (0.021)	0.141*** (0.025)	0.093*** (0.023)	0.080*** (0.018)
Location controls ^A	Yes	Yes	Yes	Yes
A-level score (cubic function)		Yes	Yes	Yes
Gender			Yes	Yes
Parental background			Yes	Yes
Other socio-economic characteristics ^A			Yes	Yes
Subject of graduation				Yes
Grade				Yes
Adjusted R ²	0.20	0.22	0.29	0.39
Observations		6986		
Continuous measure	(1)	(2)	(3)	(4)
Normalised Quality score	0.077*** (0.007)	0.055*** (0.009)	0.041*** (0.009)	0.034*** (0.007)
Normalised Quality score ²	0.009 (0.005)	0.003 (0.005)	0.007 (0.005)	0.009* (0.005)
Location controls ^A	Yes	Yes	Yes	Yes
A-level score (cubic function)		Yes	Yes	Yes
Gender			Yes	Yes
Parental background			Yes	Yes
Other socio-economic characteristics ^A			Yes	Yes
Subject of graduation				Yes
Grade				Yes
Adjusted R ²	0.20	0.22	0.28	0.38
Observations		6986		

Note: standard errors reported into brackets adjusted for clustering at the institution level. ***, **, * denote statistical significance at the 1%, 5% and 10% statistical level respectively.

^A Location controls are a set of 126 postcodes to account for local labour market characteristics. The other socio-economic characteristics are age, ethnicity, disability status, school type, whether expect to live at home when studying and tuition fee status, dummies for region of origin and region of institution.

Table 3: Propensity Score Matching Estimates of University Quality on Earnings

Treatment	Q4-Q3	Q3-Q2	Q2-Q1
Ln salary			
Treated	10.15	9.97	9.91
Control	10.00	9.91	9.93
Propensity score			
Matched treated	10.15	9.97	9.91
Matched control	10.09	9.93	9.91
ATT	0.053*** (0.018)	0.036* (0.021)	0.008 (0.021)
% treated matched	99%	94%	99%
% control used for matching ^a	86%	74%	89%
Nbr of controls accounting for 50% of match – Total number of controls used ^a	47/1526	35/1485	26/1857
Observations	3,326	3,682	3,660

Note: The propensity scores are estimated by a probit with the following covariates: an indicator of A-level score missing, a cubic in A-level score, gender, parental occupation categories, subject of degree, tuition fees status, accommodation status, disability, type of school attended, ethnicity and a set of dummies for age on graduation, and region of residence.

The matching estimator is based on Epanechnikov kernel matching with a bandwidth of 0.01

^a These statistics are based on a nearest neighbour matching with a caliper of 0.01 which are available upon request.

***, **, * denote statistical significance at the 1%, 5% and 10% statistical level respectively.

Table 4: Balancing of the GPS covariates – t-statistics of mean difference

	Unconditional			Conditional on GPS			Imai Test	
	Tier 1	Tier 2	Tier 3	Tier 1	Tier 2	Tier 3	Uncond.	Cond.
Male	4.56	4.94	0.14	0.53	1.38	0.58	3.66	0.18
A-level none	22.93	1.84	20.85	1.27	0.22	1.30	25.12	1.32
A-level]0,2]	6.42	1.50	4.88	0.26	0.13	0.39	6.75	0.84
A-level]2,4]	7.56	1.00	6.47	0.62	0.17	0.92	7.92	0.92
A-level]4,6]	4.71	2.33	6.92	0.13	0.88	0.96	8.28	0.89
A-level]6,8]	10.64	1.71	8.80	0.36	0.09	1.01	10.88	0.61
A-level]8,10]	9.64	0.15	9.65	0.40	0.65	0.93	11.71	0.77
A-level]10,12]	7.66	1.53	9.08	0.10	0.71	0.93	10.42	0.66
A-level]12,14]	5.94	3.71	9.51	0.23	0.97	0.98	10.06	1.15
A-level]14,16]	5.62	0.84	6.42	0.43	0.00	0.48	7.44	0.60
A-level]16,18]	0.32	4.95	5.10	1.00	0.41	0.13	4.13	2.57
A-level]18,20]	6.30	8.22	1.57	0.75	0.90	0.30	1.90	3.02
A-level]20,22]	9.34	4.83	4.80	0.09	0.27	0.18	6.18	3.32
A-level]22,24]	10.87	3.25	7.91	0.88	0.09	0.83	9.28	2.65
A-level]24,26]	12.06	3.04	15.51	1.07	0.96	0.16	14.16	1.12
A-level]26,28]	15.22	5.09	21.05	1.17	0.72	0.66	20.96	0.41
A-level]28,30]	18.23	13.56	33.86	1.17	1.03	2.07	40.74	10.16
Parental occupation								
gp1	7.97	0.61	7.49	0.68	0.19	0.56	9.04	0.36
gp2	13.60	3.61	17.52	1.01	1.13	2.08	19.20	0.77
gp3	0.10	1.22	1.07	0.65	0.61	0.01	0.84	0.06
gp4	0.91	0.61	1.51	0.39	0.08	0.85	0.42	0.38

Continues on next page

gp5	3.62	0.63	4.23	0.10	0.07	0.84	4.90	1.60
gp6	0.08	2.19	2.18	0.25	0.70	0.39	0.01	0.24
gp7	1.58	2.01	3.53	0.18	0.49	1.01	3.07	0.82
gp8	3.84	1.25	2.64	1.32	0.67	0.29	2.23	0.58
gp9	0.26	0.61	0.85	0.02	0.13	0.24	0.37	0.16
gp10	15.44	3.51	19.01	0.26	1.02	1.00	22.05	1.38
Disability status	1.46	5.15	3.50	0.75	1.31	1.14	2.83	1.40
School - FE institution	14.03	2.77	11.26	0.89	0.00	0.20	16.16	2.28
School - HE institution	2.91	1.45	4.31	0.06	0.61	0.78	3.89	0.27
School – Independent	14.15	6.61	21.40	1.32	0.87	2.64	24.85	3.62
School - Unknown	6.85	7.46	14.19	0.12	3.01	2.48	14.61	4.27
School - State school	9.04	1.36	10.47	0.32	0.92	0.42	12.10	0.25
White	7.75	5.55	2.46	0.82	1.67	1.18	3.72	3.18
Black	6.20	2.16	4.07	0.62	0.42	0.23	5.68	1.19
Indian	4.76	3.62	1.32	0.85	1.17	0.74	2.31	1.29
Pakistani/Bangladeshi	2.83	0.74	2.12	0.38	0.05	0.01	2.62	0.45
Chinese	0.33	1.76	2.01	0.44	0.73	0.55	2.45	0.95
Mixed	0.48	0.15	0.34	0.17	0.18	0.11	0.41	0.08
Other	3.19	2.95	0.38	0.05	0.84	0.98	0.58	2.95
Age <18	2.98	0.39	3.44	0.33	0.29	0.70	4.93	0.85
Age[18,19]	4.49	0.85	3.71	0.12	0.39	0.52	4.58	0.86
Age [19-20]	0.49	0.94	0.41	0.12	0.39	0.99	0.22	0.99
Age[20,21]	2.56	2.17	0.50	0.30	0.00	0.38	1.43	1.56
Age[21,22]	0.29	0.47	0.75	0.46	0.14	1.41	0.01	0.98
Age[22,23]	4.13	0.31	4.41	0.29	0.33	0.65	4.08	0.10

Continues on next page

Age[23,24]	1.43	2.21	3.56	0.18	0.50	0.00	4.15	0.20
Age>24	6.84	1.94	4.93	0.92	0.32	0.44	7.28	0.24
Fee status	0.23	1.80	1.50	0.88	0.47	0.55	0.05	1.23
Subject of studies								
Medicine	10.35	6.68	17.50	1.17	1.82	2.15	17.66	2.46
Sub. Allied to Med.	0.53	0.73	1.24	0.46	0.03	0.16	2.75	2.13
Biology	3.09	1.13	2.04	0.30	0.10	0.11	3.70	1.76
Physics	4.86	1.53	3.46	0.63	0.05	0.01	3.19	1.35
Math	7.32	2.77	10.27	0.87	0.78	1.12	10.02	0.26
Engineering	3.05	1.08	4.13	0.73	0.66	0.70	5.38	0.32
Architecture	1.95	1.75	3.64	0.06	0.11	0.57	3.56	1.64
Social studies	2.85	3.97	0.95	0.57	0.48	0.82	0.51	1.51
Law	4.49	1.91	6.43	0.31	0.70	0.02	7.78	0.68
Business and admin.	11.43	4.15	15.49	0.16	1.68	1.41	17.02	0.77
Communication	7.52	2.32	5.21	0.92	0.53	0.47	7.56	1.80
Linguistic	6.23	0.13	6.26	1.01	0.14	0.01	6.33	0.19
Literature	7.55	1.85	9.62	1.08	0.37	0.81	9.60	0.10
History/Philosophy	7.53	2.19	9.88	0.79	0.63	0.51	10.74	0.50
Arts	8.78	1.96	10.62	0.25	0.80	1.12	12.43	3.07
Education	6.91	6.50	0.75	1.62	1.18	1.23	3.58	0.39
Other	1.50	0.72	2.18	0.19	0.51	0.41	3.47	0.23
Sport sciences	10.66	3.89	6.71	1.34	0.57	0.48	9.08	2.20
Psychology	0.91	0.37	1.27	0.14	0.15	0.01	2.23	0.62
IT	4.65	3.08	7.62	0.27	0.75	0.94	6.44	0.68
Economics	7.02	0.21	7.47	1.03	0.05	0.42	8.49	0.11
Finance/Accounting	3.18	2.74	5.81	0.26	1.38	1.40	6.01	0.26

Continues on next page

Mixed no science	3.21	1.71	1.58	0.84	0.63	0.42	2.30	1.12
Mixed half science	2.08	0.43	2.50	1.12	0.30	1.36	0.40	0.34
Mixed 100% science	4.66	0.72	5.50	0.68	0.37	1.15	7.14	1.12
Region of residence								
UK not specified	1.70	0.10	1.86	0.61	0.07	0.07	0.93	0.42
Channel Island	1.52	2.78	1.19	0.07	0.35	0.39	0.74	0.54
Isle of Man	0.26	1.96	2.16	0.26	0.35	0.57	1.40	0.83
England not specified	5.68	3.23	2.56	1.39	1.21	0.74	3.39	1.35
Wales	1.37	1.19	0.23	0.00	0.27	0.04	0.48	1.07
Scotland	13.03	0.58	13.99	1.44	0.65	2.30	10.78	3.47
N. Ireland	9.69	17.97	7.67	2.02	5.26	1.17	1.61	4.39
East Anglia	1.47	0.19	1.67	0.40	0.13	0.07	2.35	0.29
East Midlands	6.29	1.90	4.46	0.77	0.33	0.91	3.66	1.58
greater London	1.04	3.64	2.44	0.54	0.94	1.15	2.44	3.07
rest of the world	4.22	0.04	4.23	0.18	0.25	0.57	4.25	3.07
North	5.81	7.35	1.13	0.58	1.60	1.11	1.73	0.57
North West	4.05	0.01	4.06	0.20	0.24	0.07	5.18	0.64
South East	3.66	1.99	1.78	0.47	0.60	0.76	4.12	0.62
South West	2.20	1.67	0.61	0.35	0.45	0.14	0.53	0.09
West Midlands	2.51	2.53	0.11	0.45	0.65	0.04	3.03	1.29
Yorkshire	7.68	3.86	3.97	1.09	0.84	0.71	5.46	0.69
Region of Institution								
North East	2.00	13.88	10.99	0.79	4.44	2.30	2.73	1.93
Yorkshire & Humber.	5.11	3.56	8.57	0.01	0.89	3.24	6.42	0.54
North West	5.90	0.23	6.13	0.11	0.30	0.38	10.72	1.53
East Midlands	5.15	2.29	2.96	0.42	0.44	1.58	2.11	1.77

Continues on next page

Eastern	3.60	3.84	0.05	2.51	0.60	0.17	10.12	5.22
South East	5.48	6.17	0.42	1.30	1.54	0.31	1.98	4.16
London	4.60	9.14	4.02	1.38	2.98	3.65	4.97	9.22
South West	0.39	5.84	6.04	0.09	1.26	1.34	5.24	3.71
West Midlands	4.89	7.41	2.11	0.85	1.94	0.87	4.34	0.80
Scotland	15.16	1.56	17.22	1.52	0.71	2.48	13.17	3.87
Wales	5.29	2.56	2.90	1.27	0.48	0.57	0.05	3.80
Northern Ireland	10.69	22.02	10.63	2.04	5.38	1.29	0.78	5.08

Note: Column 1 to 3 reports the value of t-test of mean differences of each covariate between observation in tier i and observations in the other two tiers of the institutional quality distribution. Column 4 to 6 reports the t-tests after blocking on GPS using 5 blocks. The mean differences are computed within each block and then the weighted average over the 5 blocks is reported for each tier. Column 7 reports t-test on the coefficient of treatment in a regression of covariate k . Column 8 reports the same t-test when conditioning on the predicted treatment value.

Table 5: Robustness Checks

	Excluding institutions with interview process	Female	Male	Quality exclude graduate prospect	Quality: Research, expenditure , student staff ratio
Discrete measure					
OLS estimates					
Quality 2	0.026* (0.014)	0.020 (0.016)	0.028 (0.020)	0.022 (0.019)	0.025* (0.014)
Quality 3	0.044*** (0.016)	0.049** (0.020)	0.036 (0.022)	0.042** (0.015)	0.048** (0.015)
Quality 4	0.065*** (0.019)	0.074*** (0.020)	0.063** (0.029)	0.052*** (0.019)	0.056*** (0.019)
Propensity score matching					
Quality 2 vs Quality 1	0.006 (0.021)	0.013 (0.022)	-0.023 (0.044)	0.022 (0.025)	0.016 (0.021)
Quality 3 vs Quality 2	0.044** (0.020)	0.051** (0.018)	-0.019 (0.037)	0.023 (0.020)	0.030** (0.015)
Quality 4 vs Quality 3	0.039** (0.019)	0.033 (0.023)	0.026 (0.031)	0.047** (0.018)	0.033** (0.019)
Continuous measure					
OLS estimates					
Quality score	0.032*** (0.007)	0.034*** (0.009)	0.025** (0.012)	0.028*** (0.008)	0.025*** (0.008)
Quality score square	0.010* (0.006)	0.002 (0.005)	0.013* (0.008)	0.008* (0.005)	0.005 (0.005)
Continuous Treatment Matching					
E[W/T=1]	9.942 (0.009)	9.890 (0.036)	9.956 (0.015)	9.892 (0.052)	9.915 (0.036)
E[W/T=2]	9.924 (0.006)	9.892 (0.023)	9.953 (0.004)	9.902 (0.022)	9.918 (0.015)
E[W/T=3]	9.976 (0.023)	9.964 (0.030)	10.024 (0.004)	9.993 (0.008)	9.984 (0.014)
E[W/T=4]	10.044 (0.032)	10.025 (0.040)	10.099 (0.011)	10.064 (0.009)	10.086 (0.011)

Note: OLS estimates are based on the full specification details of which are available in Table 2.

Propensity score matching is estimated using the full set of controls, details of which are found in Table 3

Continuous Treatment Matching is based on a propensity score estimated with the full set of parameters as explained in Figure 5. Standard errors are based on bootstrapping with 500 replications.

***, **, * denote statistical significance at the 1%, 5% and 10% statistical level respectively.

Annex 1:

Table A1: Sample Selection:

Selection criteria – applied incrementally	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
Measure of institution quality	8,500
Earnings non missing	7,508
FT employee	7,007
Earnings < £60,000	6986

Note: Longitudinal Destination of Leavers from Higher Education (2002/03)

Table A2: Distribution of Subject by Institutional Quality

Subject	Q1	Q2	Q3	Q4	Total
Medicine and Dentistry	0.00	0.00	2.83	9.52	3.04
Sub. allied to Medicine	7.12	7.39	5.56	4.96	6.26
Biology, vet, agri.	2.83	4.98	6.10	6.03	4.97
Physics	3.46	5.17	5.69	6.43	5.18
Mathematics	0.80	0.82	3.81	5.49	2.71
Engineering and Tech.	5.60	6.76	7.71	9.02	7.26
Architecture and Planning	1.47	3.84	1.12	0.97	1.85
Social Studies	4.58	6.27	7.76	5.60	6.05
Law	2.79	2.08	6.06	5.81	4.18
Business and administration	18.31	19.53	6.81	3.05	11.97
Communication	4.98	1.90	1.15	0.51	2.15
Linguistic and Classics	1.66	1.92	5.13	5.07	3.43
Language and literature	0.44	0.94	2.60	6.07	2.49
History and Philosophy	1.80	0.95	6.17	7.97	4.20
Creative Arts	10.72	8.22	4.37	1.30	6.19
Education	3.95	1.76	1.02	1.66	2.11
Sport science	4.62	1.35	1.28	0.14	1.87
psychology	3.72	3.27	3.62	2.23	3.22
IT	8.32	7.82	5.81	3.20	6.31
Economics	0.23	0.55	2.48	3.91	1.78
Finance & Accounting	2.55	2.55	1.71	0.17	1.75
mixed no science	3.94	2.65	3.38	2.90	3.22
Mixed 45-55 science	4.74	7.19	5.98	5.14	5.76
Mixed 100% science	0.41	1.10	1.40	2.65	1.38
Other	0.98	0.99	0.45	0.23	0.67
Observations	1783	1732	1758	1713	6986

Note: for each column, a cell reports the fraction of individuals in that subject. The total row reports the number of individuals.