How Robust is the Evidence on the Returns to College Choice? Results Using Swedish Administrative Data

Kent Eliasson[*]
Department of Economics, Umeå University, and National Institute for Working Life
SE-831 40 Östersund, Sweden
kent.eliasson@arbetslivsinstitutet.se

Abstract

We estimate the causal effect on earnings of graduating from old universities rather than new universities/university colleges. The study is based on Swedish administrative data that is comparatively rich in terms of school grades, parental characteristics and other attributes. Despite the more favorable conditions at old universities in terms of factors related to college quality, we find no significant difference in estimated earnings between graduates from the two groups of colleges. This finding holds for male and female sub-samples covering all majors, as well as male and female sub-samples covering two broad fields of education. The results are robust with regard to different methods of propensity score matching and regression adjustment. The results furthermore indicate little sensitivity with regard to the empirical support in the data and alternative model specifications.

Keywords: College choice; earnings; propensity score matching
JEL classification: A22; C14; I21; J31

[*] I am grateful to Roger Axelsson, Xavier de Luna, Olle Westerlund and seminar participants at the Department of Economics, Umeå University, for valuable comments. I also would like to thank the Swedish Institute for Growth Policy Studies for providing access to data.

1. Introduction

During the last few years, there has been an increased interest in examining the relationship between college choice or college quality and labor market outcomes. From the perspective of society, the question is whether graduates from different colleges differ in their productivity in the labor market. For the individual about to invest in higher education, the question is whether his/her college choice can affect employment opportunities and earnings.

The international literature focusing on labor market effects of college quality includes Black et al. (1995, 1997, 2005), Datcher Loury and Garman (1995), Behrman et al. (1996), Brewer and Ehrenberg (1996), Brewer et al. (1999), Monks (2000), Berg Dale and Krueger (2002), Chevalier and Conlon (2003), Black and Smith (2004, 2006) and Zhang (2006). The overall conclusion from this research is that college quality matter for later earnings. Gartell and Regnér (2002, 2005), Lindahl and Regnér (2005), Eliasson (2006) and Lundin (2006) are examples of recent studies using Swedish data to estimate the labor market effects of college choice. The results in these papers range from rather large to small or insignificant estimated earnings premiums of college choice.[1]

The principal econometric problem in estimating the effect of college quality or college choice on labor market outcomes follows from the non-random nature of college selection. Better students sort into more selective colleges. The standard approach in the literature has been to rely on what Heckman and Robb (1985) refer to as selection on observables to identify the effect of college choice in the presence of non-random selection of students into colleges of different types. Under this assumption, conditioning on a sufficiently rich set of observable characteristics of students removes bias resulting from non-random selection into colleges.

There are two main methods for implementing the selection on observables strategy: regression and matching. Until recently, the literature has been dominated by the former. Black and Smith (2004) discuss two related drawbacks of the conventional linear regression approach. The first weakness is that the linearity assumption can hide lack of support in the data. To illustrate the problem, consider the case where high ability individuals only graduate from high quality colleges and low ability individuals only graduate from low quality colleges. Under these circumstances, it is impossible to identify the effect of college quality on earnings without making arbitrary assumptions about the functional form of the relationship between earnings, college quality and

[1] See Eliasson (2006) for a brief summary of the findings in the cited papers.

ability. While matching estimators typically handle the support problem by dropping observations lacking sufficient support, conventional regression estimators instead achieve comparability by imposing linearity and extrapolating over regions of no support. This is the second drawback of the linear regression approach. Given the typical lack of firm theoretical guidance concerning exact model specification, the reliance on possibly incorrect functional form assumptions can generate considerable extrapolation bias.[2] On the other hand, regression is more efficient, implying smaller standard errors if the functional form assumptions hold.

The purpose of this study is to estimate the causal effect on earnings of graduating from different colleges. To assess the robustness of the results, we compare matching estimates and linear regression estimates of the earnings effects of college choice. This will provide some indications of the extent to which the estimated effects are sensitive with regard to possible incorrect functional form assumptions and support problems. To check the robustness of the matching estimates, we further report results from sensitivity analyses in three specific respects. The first concerns the choice of matching method. The second relates to the thinness of the empirical support in the data (i.e. the number of comparable observations in the treatment group and the comparison group). The third is related to the selection of covariates to be used in the estimations. The paper is inspired by Black and Smith (2004), who use a similar approach when evaluating the evidence on the effects of college quality in the United States.

For simplicity, we consider a binary treatment case and estimate the earnings premium of graduating from *old universities* rather than *new universities/university colleges*.[3] Although we do not focus on the labor market effects of college quality as such, there are important differences between the two groups in terms of factors likely to be related to college quality. The formal qualifications of teachers are one example. The percentage of faculty with doctoral degrees at old universities is about 77 percent as compared to 44 percent at new universities/university colleges.[4] The two groups also differ with regard to academic tradition. The majority of the old universities have been around for centuries, in some cases since the late middle ages. The new universities/university colleges are of a much more recent date; most of them were established only a few decades ago.[5]

A feature that makes the Swedish case particularly interesting is the availability of comparatively rich and high quality administrative data in combination with a fairly

---

[2] King and Zeng (2006) provide an interesting discussion and illustration of the problems of extrapolation bias in causal analysis.
[3] Eliasson (2006) presents results based on a more detailed college classification.
[4] These figures refer to averages for the period 1995−1999. Source: Statistics Sweden.
[5] See Appendix A for additional details of the college grouping.

straightforward institutional setting governing college selection. Our data consist of six cohorts of Swedes born in the years 1969−1974, who have completed at least a three-year college degree during the period 1994/95−1998/99, and who received positive earnings in 2003. The data set is unusually rich in terms of school grades, family background characteristics and other attributes. The admission procedure for higher education in Sweden is furthermore rather transparent and to a large extent based on observable characteristics. These circumstances provide some credibility for using selection on observables and matching to identify the earnings effect of college choice in the presence of non-random sorting of students into different colleges.

The overall conclusion from the analysis is that we cannot find any significant differences in earnings between graduates from the two groups of colleges. This holds for male and female sub-samples covering all majors, as well as male and female sub-samples covering two broad fields of education. The results are robust with regard to different methods of propensity score matching and regression adjustment. Furthermore, the results indicate little sensitivity with regard to the empirical support in the data and alternative specifications of the propensity scores. In effect, this means that the unconditional earnings premium of about 8−15 percent (depending on the sub-sample) of graduating from old universities, disappears when we compare comparable individuals using different types of propensity score matching methods and linear regression.

The paper proceeds as follows. Section 2 briefly outlines the institutional setting for higher education in Sweden. Section 3 presents the data available for the study. The econometric approach is sketched in section 4. Section 5 reports the different matching estimates and linear regression estimates of the earnings effects of college choice. Section 6 considers the sensitivity of the results with regard to the empirical support in the data. Section 7 checks robustness with regard to alternative model specifications. Section 8 summarizes the findings and provides some final remarks.

2. Higher education in Sweden

Higher education in Sweden is provided by universities and university colleges.[6] The principal difference between the two is that university colleges are generally not allowed to offer postgraduate education and to award postgraduate degrees.

[6] See Öckert and Regnér (2000) for a more detailed description of the institutional setting for higher education in Sweden.

Contrary to colleges in the United States and many European countries, there are no tuition fees at Swedish universities and university colleges. The higher education institutions are instead primarily funded by the government. They receive funding for undergraduate education based on the number of students enrolled and student performance. They also receive funding for postgraduate education and research. In addition, the government offers financial support to all students, independent of social background and the parents' financial circumstances. Currently, the study allowance is about SEK 7,000 (875 USD) per month, two thirds of which are loans and one third grants.

Historically, admission to higher education in Sweden has been unrestricted. With the 1977 Higher Education Act, the government decided that admission should be restricted and that one administrative authority, the National Swedish Board of Universities and Colleges, should handle admission to all universities and university colleges according to standardized rules of eligibility and admission. The admission requirements have changed somewhat over time, but there are no major differences between the current requirements as described below and those originally formulated in the 1977 Higher Education Act.

To be admitted to a program, a student has to attain a general admission requirement, which basically means having completed an upper secondary education in Sweden or abroad. This requirement can also be met by work experience in combination with adequate knowledge in Swedish and English. In addition, most programs have specific admission requirements, such as sufficient knowledge in key subjects. Since the number of applicants for a particular program typically exceeds the number of places available, fulfilling the general and specific admission requirements is not sufficient to be admitted. Applicants are then ranked, primarily on basis of their upper secondary school grade point average (GPA) and scores on the Swedish Scholastic Aptitude Test (SAT). Those applicants who have been ranked highest are admitted to the program.[7]

The lack of tuition fees in combination with financial support for all students imply that the financial situation of the students is not likely to directly affect college choice in Sweden. The admission procedure for higher education is furthermore fairly transparent and primarily based on observable qualifications, mainly upper secondary school GPA and SAT scores. Hence, Swedish colleges are not allowed to choose freely among eligible applicants. This differs from the situation in, for example, the United States, where unobservable factors play a larger role for college admission decisions.[8]

[7] In practice, the share of places allocated on the basis of upper secondary school GPA dominates.
[8] See Berg Dale and Krueger (2002) for a discussion of observable and unobservable college admission characteristics in the United States.

Altogether, the intuitional setting governing college admission in Sweden contributes to the plausibility of selection on observables and matching as a reasonable identification strategy.

3. Data

This study is based on a data set that has been constructed from a number of administrative registers kept by Statistics Sweden.[9] The data set includes six cohorts of Swedes born in the years 1969−1974, who have completed at least a three-year college degree during the period 1994/95−1998/99, and who received positive earnings in 2003.[10,11] Each person fulfilling these requirements is classified as having a degree from an old university or a new university/university college, depending on from which type of college he or she graduated.[12] The period from college graduation to measured earnings is long enough for most individuals to have become well established in the labor market. The average potential labor market experience after college graduation is almost identical in the two groups; approximately seven years.[13] Given the strong empirical evidence on the importance of experience for earnings in early working life, this is a desirable outcome of the sampling procedure.

The identification strategy in the paper requires that we observe all variables influencing both the treatment and the outcome. On the basis of this condition and the guidance of economic theory and previous empirical research, we have put together a data set including (1) basic individual characteristics such as age, gender, country of birth and region of residence; (2) grade point average and study program in upper

[9] The following registers have been used: the Register of the Total Population, the Register of the Population's Education, the Register of Universities and University Colleges, the Register of Grades from Upper Secondary School, the Register of Income Statements and the Register of Income, Taxes and Allowances.

[10] College graduates from the Swedish University of Agricultural Sciences, artistic colleges and colleges run by the county councils are excluded.

[11] Since the sample is restricted to students who have completed at least a three-year college degree, there is some potential risk for dropout bias in the presented results. This could be the case if dropouts vary systematically between the college groups. Unfortunately, the data does not allow us to pursue this issue. See Öckert (2001) for a general discussion and analyses of potential problems with dropout bias in a Swedish context.

[12] 12 percent of the individuals in the data set completed more than one college degree. In this case, the degree corresponding to the highest number of credits is selected in a first stage and, if necessary, the most recent one is chosen in a second stage. Note that some students may begin their studies at one college and graduate from another. A student's college classification (old/new) is always based on the type of college he or she graduated from.

[13] The exact figures are 6.7 years for graduates from old universities and 6.8 years for graduates from new universities/university colleges.

secondary school;[14] (3) parental characteristics such as age, country of birth, level of education and earnings of the mother and the father; (4) neighborhood attributes such as the level of education and average earnings in the parish of residence; (5) information on the identity of the degree awarding college, field/major and number of credits of the degree.[15] All family background and neighborhood characteristics, as well as information on the individual's region of residence, refer to the situation at age seventeen. This is roughly a year prior to the earliest possible age of college enrollment.

The dependent variable in the analysis is the log of total annual earnings from employment and self-employment in 2003. Annual earnings are a function of both hourly wages and number of hours worked during a year. To enhance the comparability with previous Swedish studies, we will primarily report results for individuals with total annual earnings above SEK 100,000 (12,500 USD). Antelius and Björklund (2000) show that with this restriction, the estimated effects of education on annual earnings are similar to those obtained using hourly wages. However, it is important to note that the effect of unemployment and labor supply decisions on annual earnings is not necessarily exogenous with regard to college choice (and college quality). Therefore, we will briefly comment on results based on a more moderate restriction, allowing for all positive earnings. From an economic perspective, it is difficult to argue that one of the two applied restrictions is necessarily better than the other. With the restriction of SEK 100,000, the focus is primarily on the productivity of individuals who are employed. With the more moderate restriction, the focus is both on employment opportunities and worker productivity.

In total, there are 53,342 individuals meeting the conditions: born in the years 1969−1974; completed at least a three-year college degree during the period 1994/95−1998/99; and received total annual earnings from employment and self-employment above SEK 100,000 in 2003. Grades in upper secondary school are available for 89.4 percent, parental characteristics are available for 93.7 percent and neighborhood attributes are available for 98.7 percent. In total, 16.0 percent of the observations have missing values for at least one variable. After deleting observations with missing values, the data set is reduced to 44,807 individuals.[16]

[14] Eliasson (2006) uses several alternative sets of school grades as indicators of unobserved true ability, including grades in specific subjects as well as grade point average at both the compulsory school and the upper secondary school level. The combination of grade point average and study program in upper secondary school performs best in terms of explaining post-college graduation earnings.

[15] We refer to the cited papers in the introduction for theoretical and empirical motivations for the variables used in the analysis. Willis (1986) and Card (1999) are two excellent survey articles in the field.

[16] A comparison between the initial data set of 53,342 individuals and the final data set of 44,807 individuals with complete information, reveals a bias towards individuals born in Sweden in the latter (98.6 percent compared to 84.1 percent). This is expected, since information on earlier school achievements and family background is more likely to be missing for immigrants.

Table 1. Sample means by college group (total number of observations is 44,807)

	Old universities		New universities/ university colleges	
	Men	Women	Men	Women
Outcome variable				
Total annual earnings 2003 (SEK 1,000)	396.7	281.4	347.5	243.8
Individual attributes				
Age	31.8	31.5	31.5	31.0
Upper secondary school grade point average	3.95	3.99	3.58	3.65
Parental characteristics				
Dad college education, %	45.8	43.0	33.8	30.2
Dad total annual earnings (SEK 1,000)	320.2	310.9	269.5	258.7
Mom college education, %	49.7	49.1	37.9	36.4
Mom total annual earnings (SEK 1,000)	172.6	175.9	157.1	157.5
Neighborhood characteristics				
3-year college education or higher, %	11.4	11.4	8.4	8.3
Average total annual earnings (SEK 1,000)	155.0	154.1	144.4	144.8
Degree level (semesters)	8.2	7.7	7.3	7.0
Number of observations	12,028	11,297	9,445	12,037

In the empirical sections to follow, the data set is further split into six sub-samples. These include all male and female college graduates (irrespective of major), and male and female college graduates in two specific fields of education, namely law and social sciences and technology. The main arguments for focusing specifically on these fields of education are that they are well represented in both college groups and that they include types of education typically demanded in both the private sector and the public sector. Law and social sciences is dominated by graduates in business and economics, whereas technology is dominated by graduates in engineering.

Table 1 present basic descriptive statistics by college group for the sub-samples including all male and female college graduates.[17] Three main findings emerge from the sample characteristics. First, there are large earnings differentials between the two college groups. College graduates from old universities have, on average, about 15 percent higher annual earnings than graduates from new universities/university colleges. Second, the upper secondary school grade point average is higher for college graduates from old universities. Third, college graduates from old universities generally come from a more favorable background, both in terms of parental characteristics and

[17] Complete descriptive statistics are available from the author.

Table 2. Variable description

Outcome variable	
Log of total annual earnings 2003	Log of total annual earnings from employment and self-employment in 2003 (2003 SEK).
Individual attributes	
Age	Age in 2003.
Born in Sweden	Dummy variable indicating born in Sweden.
Region of residence	A set of 5 dummy variables indicating region of residence (Stockholm county, Uppsala county, Skåne county, Västra Götaland county, other counties).
Upper secondary school grade point average	Grade point average based on grades in all courses (about 18) according to a five point number scale (1−5). Quadratic is used. Complemented by a set of 6 dummy variables indicating study program.
Parental characteristics	
Dad/mom age	
Dad/mom born in Sweden	Dummy variable indicating born in Sweden.
Dad/mom level of education	A set of 4 dummy variables indicating level of education (primary and secondary, upper secondary, college, graduate).
Dad/mom total annual earnings from employment (SEK 1,000)	Total annual earnings from employment (2003 SEK).
Neighborhood characteristics	
3-year college education or higher	Percent of working age population (20−64) in the parish of residence with ≥ 3-year college education or graduate education.
Total annual earnings (SEK 1,000)	Average total annual earnings from employment of working age population (20−64) in the parish of residence (2003 SEK).
College education characteristics	
Degree level	Length of college degree in semesters.
Field/major	A set of 7 dummy variables indicating college field/major.

Note: All parental and neighborhood characteristics, as well as information on the individual's region of residence, refer to the situation at age seventeen.

neighborhood attributes. Note that the differences between the two college groups hold for male as well as female college graduates. Table 2 provides a detailed description of the variables used in the analysis.

4. Econometric strategy and the parameter of interest

The main purpose of this paper is to estimate the causal effect on earnings of college choice using propensity score matching methods. In this section, we give a brief presentation of how matching solves the evaluation problem and discuss some of the practical issues faced when implementing propensity score matching.[18]

A specific feature of our application is that all individuals receive treatment in the literal sense. Let Y_1 be the potential earnings associated with graduating from old universities and Y_0 the potential earnings associated with graduating from new universities/university colleges. Furthermore, let $D = 1$ indicate receiving treatment from an old university and $D = 0$ indicate receiving treatment from a new university/university college. Finally, let X denote a set of observed variables affecting both college selection and earnings.

The main parameter of interest is the average treatment effect on the treated, *ATT*, which can be defined as:

$$ATT = E(Y_1 - Y_0 | D = 1) = E(Y_1 | D = 1) - E(Y_0 | D = 1) \tag{1}$$

In our context, *ATT* corresponds to the average effect on earnings of graduating from old universities rather than new universities/university colleges for those persons who actually graduate from old universities.[19] The fundamental evaluation problem is that we only observe Y_1 or Y_0 for each person, but never both. $E(Y_1 | D = 1)$ can be constructed directly from the data. Missing is the information required to identify $E(Y_0 | D = 1)$, referred to as the counterfactual outcome. If college selection is non-random and we substitute the unobservable $E(Y_0 | D = 1)$ for the observable $E(Y_0 | D = 0)$ when estimating *ATT*, we end up with selection bias equal to $E(Y_0 | D = 1) - E(Y_0 | D = 0)$.

The method of matching solves the evaluation problem by assuming that, conditional on X, Y_0 is independent of D:

$$Y_0 \perp D | X \tag{2}$$

This is referred to as the conditional independence assumption (CIA). The intuition behind this crucial assumption is that it makes treatment assignment random conditional on X, which in a sense ex post reproduces the essential feature of a randomized experiment. When CIA holds, we can therefore use the earnings of graduates from new

[18] For a more detailed and technical presentation of matching methods, see e.g. Heckman et al. (1998a), Imbens (2004) and Smith and Todd (2005a).

[19] If treatment effects are heterogeneous, *ATT* will differ from the average effect of graduating from old universities for those individuals who actually graduate from new universities/university colleges (average treatment effect on the untreated, *ATU*) and from the average effect of graduating from old universities for a randomly selected person (average treatment effect, *ATE*). In Section 5, we will briefly comment on possible effect heterogeneity.

universities/university colleges as an approximation of the counterfactual outcome (what graduates from old universities would have earned had they graduated from new universities/university colleges). Formally expressed, we have $E(Y_0 | X, D = 1) = E(Y_0 | X, D = 0)$, which allows for an unbiased estimation of *ATT*.

Furthermore, matching methods rely on a common support or overlap condition, which for *ATT* can be formally stated as:

$$\Pr(D = 1 | X) < 1 \tag{3}$$

This condition prevents *X* from being a perfect predictor of treatment status. In our context, this ensures that for every *X*, there are persons graduating from both old universities and new universities/university colleges, which means that for every *X*, we will be able to construct the counterfactual outcome.

At this point, it can be interesting to reflect upon how matching methods differ from linear regression. Since conventional regression techniques also rely on an assumption of conditional mean independence, one might argue that the difference between the two is merely cosmetic. But while linear regression rests on the additional assumption that simply conditioning linearly on *X* is sufficient to remove selection bias, matching methods handle the selection problem either by non-parametric or semi-parametric techniques (depending on the particular method employed). Another important difference is that conventional regression estimates typically are obtained without ensuring that there actually exist comparable treated and non-treated observations for every *X*. The evidence in Heckman et al. (1998b), Dehejia and Wahba (1999, 2002) and Smith and Todd (2005a) suggests that avoiding functional form assumptions and imposing a common support condition can be important for reducing selection bias.

The basic idea of matching is to pair treated and non-treated observations on the basis of their observable characteristics. When *X* is of high dimension, which is the case in this application, it becomes difficult to find close matches along all dimensions of *X*. Rosenbaum and Rubin (1983) show that if matching on *X* is valid, so is matching on the conditional probability of receiving treatment, $\Pr(D = 1 | X)$, referred to as the propensity score. The benefit of using the propensity score is that it reduces the dimensionality of the matching problem, thus allowing us to match on a scalar variable rather than in a multidimensional *X*-space.

While propensity score matching has many desirable properties, it is not (using the words of Smith and Todd, 2005a) a "magic bullet" that solves the selection problem in all situations. The method critically depends on the conditional independence assumption, which requires that all variables affecting both the treatment and the outcome are included in the matching. In this particular application, the propensity score matching approach is justified on the basis of a fairly transparent institutional setting governing college selection in combination with access to comparatively rich and high quality administrative data. But even though the conditions in this sense are favorable, it is important to be aware of the limitations of propensity score matching. A major one is that the method assumes away selection on unobservables as a source of bias.[20] This is undoubtedly a strong assumption. But given the data available and the institutional setup, we consider propensity score matching a natural starting point for estimating the causal effect on earnings of college choice.

A number of questions arise when implementing propensity score matching methods in practice.[21] A first concerns which method to use when estimating the propensity score. The discussion in e.g. Smith (1997) suggests that this choice is not too crucial. Any flexible parametric estimator, such as a logit model or a probit model, will do. In this application, we use a probit model to estimate the propensity score.

The next question concerns the selection of covariates to match on. This requires choosing a set of variables plausibly satisfying the conditional independence assumption. There exists no deterministic rule for this and the results in Heckman et al. (1997, 1998b) show that matching estimates can be very sensitive with regard to the choice of covariates used to estimate the propensity score. In this application, we use a set of conditioning variables which are commonly found in the related literature and typically argued to influence both college choice and earnings. To check the sensitivity of the results, we also experiment with alternative sets of conditioning variables. To make the estimated treatment effect clearly interpretable, it is particularly important to avoid conditioning on covariates that are determined by the treatment (Rosenbaum, 1984; Heckman et al., 1999; Imbens, 2004). In this particular setting, conditioning on post-college graduation variables (such as experience, region or sector of employment) can result in a biased estimate of the treatment effect because these variables may have been affected by the treatment, and thereby carrying part of the effect. Therefore, we restrict our set of X variables to only include covariates determined prior to college graduation.

Having decided upon the model to estimate the propensity score and the covariates to match on, the remaining question is which matching algorithm to use. Asymptotically, all matching estimators are consistent because as the sample size grows, they all end up comparing only exact matches (Smith, 2000). In finite samples, the choice of matching

[20] This comment refers to cross-sectional matching estimators, which are in focus here. There are, however, difference-in-differences matching strategies which can eliminate selection bias due to time invariant unobservable characteristics; see e.g. Heckman et al. (1997, 1998b).

[21] Caliendo and Kopeinig (2006) provide an accessible overview of the typical issues one faces.

algorithm can make a difference. All matching estimators are weighting estimators in the sense that they take a weighted average of the outcomes of the untreated observations to construct an estimate of the unobserved counterfactual for each treated observation. The key difference between the various methods is how they construct the weight, which typically involves a trade-off between bias and variance.[22] In our application, the choice of matching algorithm has essentially been guided by the data at hand.[23] One specific aspect of the data is that in all dimensions that we try to estimate earnings effects of college choice, we always have more treated observations than comparison observations (i.e. the number of graduates from old universities exceeds the number of graduates from new universities/university colleges). Another feature of the data is that the distribution of the estimated propensity scores for the treatment and the comparison group tends to be rather different. Both these circumstances call for matching with replacement. Given this in combination with the large sample available, we rely on single nearest neighbor matching with replacement as our main strategy. To check the robustness of the results, we also present estimates obtained using the Epanechnikov kernel with three widely spaced bandwidths.

5. Estimates of the effect of college choice

In this section, we present our estimates of the earnings effect of graduating from old universities rather than new universities/university colleges. We report separate estimates for the six sub-samples meeting the requirements outlined in Section 3. In all cases, the outcome variable is the log of total annual earnings from employment and self-employment in 2003. Table 2 provides a detailed description of the variables we match on/control for. As mentioned previously, we restrict our set of covariates to only include factors determined prior to college graduation. The only potentially controversial choice in this sense, is conditioning on college field/major and degree level. We cannot observe whether a student chose field of education and degree level before, simultaneously with or after college choice. However, the important thing to note is that both field of education and degree level are determined prior to college graduation and hence, can be regarded as pre-treatment variables.[24] We also know that

[22] See e.g. Smith and Todd (2005a) for a technical presentation of different weighting regimes.
[23] There are more formal data-driven methods that can assist in choosing matching algorithm, such as cross-validation; see e.g. Black and Smith (2004).
[24] Still, there has been some controversy in the literature as to whether length of college education should be regarded as an exogenous pre-treatment variable, as it might in part depend on college quality; see e.g. Black and Smith (2004).

there is considerable heterogeneity in the treatment received at the two groups of colleges in terms of majors offered and length of college education. In order to compare comparable treatments, we therefore include college field/major and degree level in our set of conditioning variables.[25]

Before turning to the actual estimates, we begin by examining the distribution of the estimated propensity scores in Figure 1. For each graph, the top histogram refers to college graduates from old universities (the $D = 1$ group), whereas the bottom histogram refers to college graduates from new universities/university colleges (the $D = 0$ group). The horizontal axis delimits intervals of the propensity score and the height (depth) of each bar on the vertical axis shows the fraction of observations with scores in the corresponding interval.

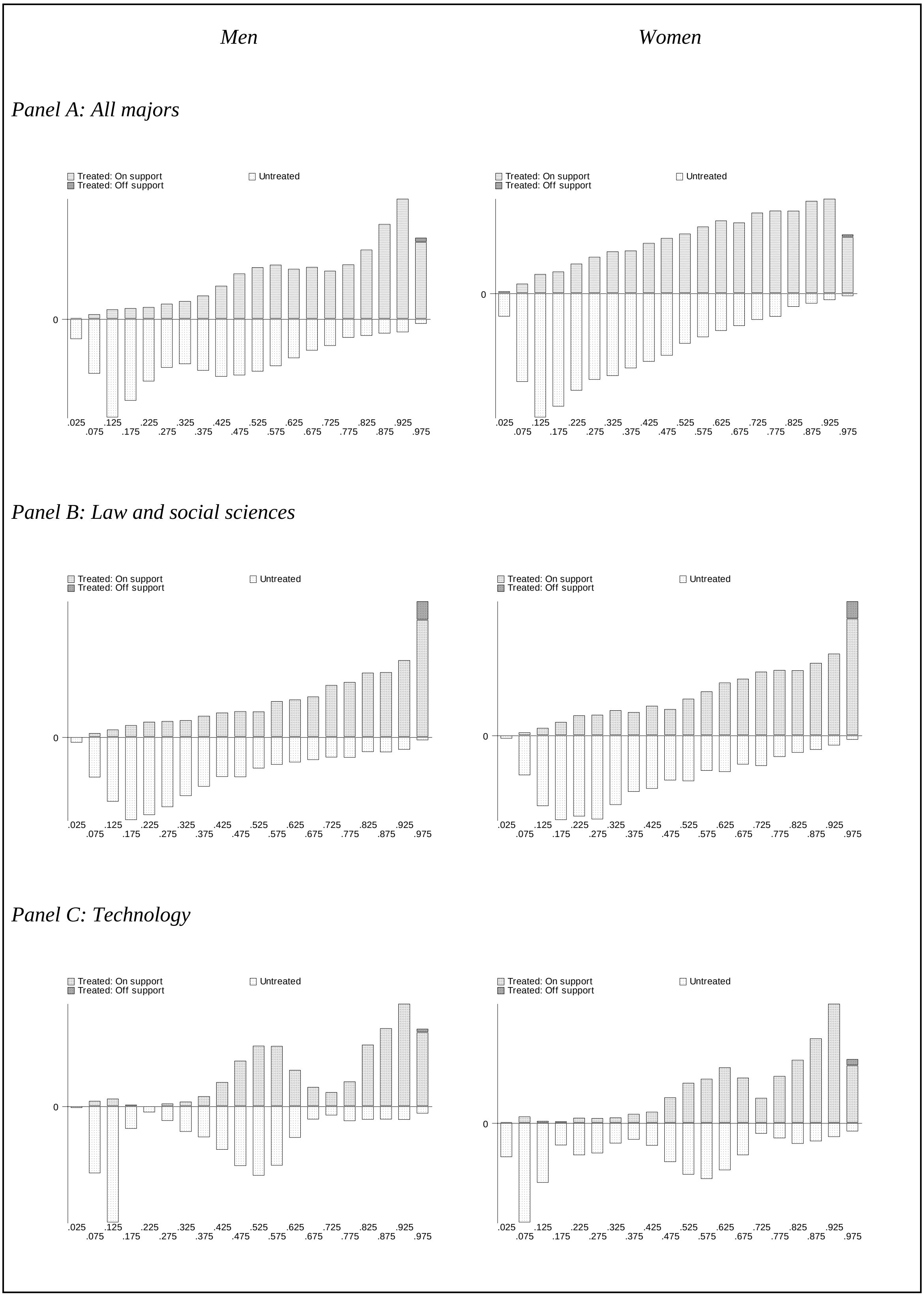
Two interesting findings emerge from Figure 1. First, apart from a few treated observations (the $D = 1$ group) with propensity scores equal to 1.0, we have support over the entire [0,1] interval for all sub-samples. Second, the distribution of the estimated propensity scores for the two groups is very different. For the $D = 0$ group, the density is clearly concentrated towards low scores, whereas the density for the $D = 1$ is distinctly concentrated towards high scores. In particular, this means that for high values of the propensity score, we only have a small number of graduates from new universities/university colleges to match against a large number of graduates from old universities. We will return to this issue in Section 6.

Table 3 presents estimates of the average treatment effect on the treated using single nearest neighbor matching with replacement. As a sensitivity analysis, the table also reports estimates based on the Epanechnikov kernel with three different bandwidths.[26] In addition, the table presents linear regression estimates of the treatment effect. The OLS estimates are based on all observations in the respective sub-samples, whereas the matching estimates only include observations on the common support. Bootstrap standard errors based on 250 replications are reported in parentheses below each matching estimate.[27]

The first line in Table 3 presents the nearest neighbor estimates. The estimated effects are generally very small. In the field of technology, the estimates suggest a modestly negative impact on earnings of graduating from old universities, both for men and women. But neither of these estimates is statistically significant at conventional levels. The effects remain very small when we switch from nearest neighbor to the

[25] The results in Lindahl and Regnér (2005) and Eliasson (2006) show that controlling for college field/major and degree level makes a large difference to the estimates.
[26] All matching estimates are obtained using PSMATCH2 for STATA, by Leuven and Sianesi (2003).
[27] Imbens (2004) notes that even though there is little formal evidence to justify bootstrapping in a matching context, the method is likely to produce valid standard errors in the propensity score case.

Figure 1. Distribution of estimated propensity scores



Note: See Table 2 and Table 3 for a description of the covariates used to estimate the propensity scores.

Table 3. Propensity score and regression estimates of the effects of college choice on annual earnings

	All majors		Law and social sciences		Technology	
	Men	Women	Men	Women	Men	Women
Nearest neighbor(1)	0.002	0.001	0.010	−0.009	−0.019	−0.020
	(0.0117)	(0.0105)	(0.0317)	(0.0232)	(0.0116)	(0.0308)
Epan(0.01)	0.003	−0.003	0.016	−0.009	−0.010	−0.013
	(0.0074)	(0.0093)	(0.0233)	(0.0208)	(0.0098)	(0.0236)
Epan(0.05)	0.003	−0.002	0.017	0.000	−0.008	−0.011
	(0.0070)	(0.0084)	(0.0195)	(0.0173)	(0.0096)	(0.0219)
Epan(0.10)	0.008	0.004	0.026	−0.001	−0.006	−0.014
	(0.0081)	(0.0077)	(0.0172)	(0.0176)	(0.0112)	(0.0210)
OLS	0.004	0.003	0.006	0.005	0.003	0.004
	(0.0053)	(0.0056)	(0.0119)	(0.0117)	(0.0074)	(0.0185)
Balancing indicators						
Mean bias before	23.5	25.3	26.8	26.3	26.4	27.0
Mean bias after	3.2	3.4	6.0	6.2	4.5	3.5
Pseudo R^2 before	0.259	0.237	0.288	0.264	0.266	0.284
Pseudo R^2 after	0.009	0.009	0.018	0.014	0.014	0.012
Untreated on support	9,445	12,037	2,639	3,051	3,931	1,052
Treated on support	11,966	11,270	3,463	3,932	5,748	1,645
Treated off support	62	27	90	83	32	16
Number of observations	21,473	23,334	6,192	7,066	9,711	2,713

Notes: The outcome variable is the log of total annual earnings from employment and self-employment in 2003. The propensity scores are estimated using a probit model and the specification includes individual attributes, quadratics in grade point average and indicators of study program in upper secondary school, parental and neighborhood characteristics, college degree level and indicators of college field/major. For the two sub-samples covering all majors, the specification includes a total of 34 covariates. The specification for the other four sub-samples includes a total of 28 covariates (excluded here are the indicators for college field/major). The specification for the OLS estimates is identical. Bootstrap standard errors based on 250 replications appear in parentheses below the matching estimates and robust standard errors are reported in parentheses below the OLS estimates. We refer to the text for a discussion of the balancing indicators. See Table 2 for a detailed description of the conditioning variables.

Epanechnikov kernel estimates. In the field of law and social sciences, the estimates indicate a modestly positive earnings premium of graduating from old universities for men as the bandwidth increases. But once more, these estimates are not statistically significant. Turning to the OLS estimates, we see that the pattern of extremely small earnings effects of college choice is verified, in this case for all six sub-samples.

The matching results reported in Table 3 are estimates of the average treatment effect on the treated (*ATT*), which in our setting corresponds to the average effect on earnings of graduating from old universities rather than new universities/university colleges for those persons who actually graduate from old universities. If impacts are heterogeneous, this effect will differ from the average effect of graduating from old universities for those individuals who actually graduate from new universities/university colleges (*ATU*) and from the average effect of graduating from old universities for a randomly selected individual (*ATE*). To check for effect heterogeneity, we have estimated *ATU* and *ATE* using single nearest neighbor matching with replacement and the same specification of the propensity scores as in Table 3. The results from this exercise (not reported) show no indications of heterogeneous impacts. The estimated parameters continue to be very small and not statistically significant in any of the six sub-samples.[28]

As mentioned previously, we follow earlier Swedish research in the field and base our reported results on individuals with total annual earnings above SEK 100,000. With this restriction on earnings, we reduce the effect of labor supply decisions and implicitly focus on the productivity of individuals who are employed. But if labor supply decisions vary systematically among graduates from the two college groups, the earnings restriction may generate biased estimates. To get an idea of the consequences of excluding individuals with a relatively weak position on the labor market, we have estimated treatment effects based on a more moderate restriction, allowing for all positive earnings. These complementary results (not reported) again show that there are no significant differences in estimated earnings between graduates from the two college groups for any of the six sub-samples.[29]

Table 3 also reports some basic information concerning the quality of the matching.[30] The general idea is to check whether there are differences in the covariates in *X* between the treatment and the comparison group after conditioning on the propensity score. If there are significant differences, the matching procedure has not been successful in terms of balancing the distribution of variables in the two groups, which calls for a re-specification of the propensity score. A problem with this approach is that there exist no formal criteria in the literature for what constitutes a successful balancing.[31] Nevertheless, we present results from two commonly applied indicators of overall covariate balancing.[32]

The first is the mean absolute standardized bias over all covariates used in the propensity score, which for the different sub-samples is around 25 percent before

[28] Complete results are available from the author.
[29] Estimates based on single nearest neighbor matching with replacement and the same specification of the propensity scores as in Table 3. Complete results are available from the author.
[30] Caliendo and Kopeinig (2006) provide a brief overview of the indicators typically used in the literature.
[31] See Smith and Todd (2005b) for a discussion.
[32] See e.g. Sianesi (2004) for a similar presentation.

matching and about 3−6 percent after matching.[33] In other words, the matching generates a reduction in mean bias by roughly a factor of four. The other indicator is the pseudo-R^2 before and after matching. This statistic indicates how well the variables in X explain the probability of receiving treatment. After matching, the pseudo-R^2 should be rather low since there should be no systematic differences in the distribution of covariates between the treatment and the comparison group. Before matching, the covariates in X explain around 25 percent of the treatment assignment in the different sub-samples. After matching, this figure drops to roughly 1 percent.

The overall conclusion from the balancing indicators is that the proposed specification of the propensity score is fairly successful in terms of balancing the distribution of covariates between the two groups.

6. Sensitivity with regard to the empirical support

Figure 1 in the previous section revealed that the distribution of the estimated propensity scores for the two groups of college graduates is very different. For the $D = 0$ group, the distribution is distinctly skewed towards low scores, whereas the distribution for the $D = 1$ group is clearly skewed towards high scores. Although the support condition does not fail in our data, the identification of the treatment effect for high values of the propensity score rely on a rather thin empirical support. Comparing the comparable in these regions of the data means using only a small number of graduates from new universities/university colleges as counterfactuals for a large number of graduates from old universities.

In this section, we follow Black and Smith (2004) and examine how sensitive the estimated effects are with regard to the thinness of the empirical support for high values of the propensity score. There are several reasons why the limited support can affect the results. With just a small number of comparison observations, it can be difficult to obtain sufficient covariate balancing for high propensity scores. Using the same comparison observation many times also keeps bias low at the cost of increased

---

[33] The standardized bias (SB) of a covariate is defined as the difference of the sample means in the treatment and the comparison group, scaled by the square root of the average of the sample variance in the two groups:

$$SB_{before} = 100\frac{\overline{X}_1 - \overline{X}_0}{\sqrt{(V_1(X) + V_0(X))/2}} \qquad SB_{after} = 100\frac{\overline{X}_{1M} - \overline{X}_{0M}}{\sqrt{(V_{1M}(X) + V_{0M}(X))/2}}$$

where X_1 (V_1) and X_0 (V_0) is the sample mean (variance) in the treatment group and the comparison group before and after matching (the latter indicated by subscript M). This measure was suggested by Rosenbaum and Rubin (1985).

Table 4. Distribution of weights for single nearest neighbor matching with replacement

Deciles of the distribution of the estimated propensity score	All majors		Law and social sciences		Technology	
	Men	Women	Men	Women	Men	Women
$P(X) < 0.1$	1.05	1.04	1.11	1.00	1.06	1.00
$0.1 \leq P(X) < 0.2$	1.09	1.14	1.19	1.19	1.06	1.00
$0.2 \leq P(X) < 0.3$	1.25	1.23	1.25	1.32	1.10	1.10
$0.3 \leq P(X) < 0.4$	1.42	1.39	1.48	1.42	1.35	1.56
$0.4 \leq P(X) < 0.5$	1.67	1.62	1.75	1.60	1.83	1.67
$0.5 \leq P(X) < 0.6$	2.07	2.01	2.11	2.05	2.13	1.83
$0.6 \leq P(X) < 0.7$	2.46	2.67	2.75	2.79	2.47	2.67
$0.7 \leq P(X) < 0.8$	3.38	3.64	3.89	3.95	2.86	4.66
$0.8 \leq P(X) < 0.9$	6.88	7.31	6.51	6.07	8.13	6.33
$0.9 \leq P(X)$	13.62	14.27	16.31	17.18	12.29	11.92
$P(X) < 1.0$	2.98	2.52	3.30	3.10	3.32	3.47

Notes: The propensity scores are estimated using a probit model. See Table 3 for the specification of the propensity scores.

variance and larger standard error of the estimated treatment effect. In addition, Black and Smith (2004) argue that the thinness of the empirical support can increase possible bias due to remaining selection on unobservables. They reach the intuitively reasonable conclusion that the effect of unobservable factors is larger at the tails of the propensity score than for scores around 0.5.

Table 4 complements Figure 1 by reporting the distribution of weights across deciles of the estimated propensity score for single nearest neighbor matching with replacement. The weights show the average frequency with which a comparison observation is used as a match in different intervals of the propensity score. The weights are fairly small in all six sub-samples up to scores around 0.7, after which a substantial increase takes place. For scores above 0.8, the average number of times a comparison observation is used as a match is between 6 and 17 times in the different sub-samples. This can be compared to an average of about 3 for the entire common support (final row in the table).

On the basis of Table 4 and Figure 1, we follow Black and Smith (2004) and present estimates based on the "thick support" region, defined as the region with an estimated propensity score in the interval $0.33 < \hat{P}(X) < 0.67$. This region is characterized by having a substantial number of observations in both the treatment group and the comparison group, which means that the average frequency with which a comparison observation is used as a match is comparatively low.

The estimated effects for the thick support region using single nearest neighbor matching with replacement are reported in Table 5. Three findings emerge from the table. First, the results seem fairly robust with regard to the thickness of the empirical

Table 5. Propensity score estimates of the effects of college choice on annual earnings for the thick support region

	All majors		Law and social sciences		Technology	
	Men	Women	Men	Women	Men	Women
Nearest neighbor(1); thick support region	0.012 (0.0111)	0.023 (0.0114)	0.030 (0.0217)	0.011 (0.0227)	−0.014 (0.0133)	0.032 (0.0313)
Balancing indicators						
Mean bias before	5.5	4.5	4.7	4.6	7.2	7.4
Mean bias after	3.4	3.1	3.2	3.6	5.1	8.1
Pseudo R^2 before	0.033	0.039	0.043	0.049	0.041	0.045
Pseudo R^2 after	0.009	0.008	0.010	0.011	0.023	0.049
Untreated on thick support	3,738	4,318	895	1,130	1,934	413
Treated on thick support	4,038	4,147	940	1,176	2,246	524
Number of observations	7,776	8,465	1,835	2,306	4,180	937

Notes: The outcome variable is the log of total annual earnings from employment and self-employment in 2003. The propensity scores are estimated using a probit model. See Table 3 for the specification of the propensity scores. Bootstrap standard errors based on 250 replications appear in parentheses below the estimates.

support (compare with Table 3). Although the estimates generally indicate a slight increase in the earnings impact of graduating from old universities, the estimated effects continue to be small and statistically insignificant. The only exception is a 2 percent positive earnings premium of graduating from old universities for women in the sub-sample covering all college majors. The fact that the estimated effects on the thick support are more or less similar to those on the entire common support is an indication of effect homogeneity over different values of the propensity score. This is what we would expect, given the similarity in the estimates of *ATT*, *ATU* and *ATE* in Section 5.

The second finding is that the thick support condition alone is almost sufficient to balance the distribution of covariates between the treatment group and the comparison group. This means that already at the outset, we compare relatively comparable individuals when focusing on the thick support region. After matching, the balancing indicators generally show slightly better overall covariate balancing as compared to the achieved balancing over the entire common support.

The third finding is that imposing the thick support condition amounts to dropping on average two thirds of the observations in the different sub-samples. The estimated effects for the thick support region thus refer to samples that, in terms of sheer size, are very different to those on the entire common support.

7. Sensitivity with regard to the specification of the propensity score

The primary assumption that justifies matching is the conditional independence assumption. The intuition behind this crucial assumption is that it makes treatment assignment random, conditional on the set of observed variables. The conditional independence assumption requires that we include all variables affecting both the treatment and the outcome in the analysis. A fundamental problem with the assumption is that it is inherently untestable (Imbens, 2004). Ideally, economic theory and previous empirical research should give clear guidance in determining which covariates that plausibly satisfy the conditional independence assumption. In reality, there exists no deterministic rule for choosing the optimal set of conditioning variables, other than comparing the resulting estimates with those from an experiment (Smith, 2000). Using experimental estimates as a benchmark, Heckman et al. (1997, 1998b) show that matching estimates can be very sensitive with regard to variable selection and that cruder variable sets generally induce larger biases.

Existing strategies for covariate selection in a propensity score matching context typically focus on two criteria: the ability of the variables to correctly predict treatment assignment and the statistical significance of the variables (see e.g. Heckman et al. 1997, 1998b). de Luna and Waernbaum (2005) suggest a procedure that not only focuses on the variables relevance for treatment assignment but also on their ability to predict outcomes. This approach has an intuitive appeal, since we know that the conditioning set must include all variables affecting *both* the treatment and the outcome for the conditional independence assumption to be satisfied. Augurzky and Schmidt (2001) employ a similar strategy in a simulation study.

Evidently, there are several methods around for selecting the appropriate set of conditioning variables. Rubin and Thomas (1996) raise the question of the appropriateness of trimming the propensity score model to achieve a more parsimonious specification. They advise against excluding a variable unless there is consensus that it is unrelated to the outcome or not a proper covariate. This brings us back to the point that the choice of variables should ultimately be made on the basis of economic theory and previous empirical findings.

Nevertheless, we will test the robustness of our results by experimenting with alternative sets of conditioning variables. We follow the two-step procedure suggested by de Luna and Waernbaum (2005). We stress that our ambition here is not to identify an optimal set of covariates. The approach is merely used to check the sensitivity of our results with regard to different specifications of the propensity scores.

The point of departure is the conditioning set *X* used in Sections 5 and 6. For the two sub-samples covering all majors, the total number of covariates available for selection is 34, and for the other four sub-samples the total number of available variables is 28 (excluded here are the indicators for college field/major). In a first step, the covariates predicting treatment assignment are identified. The second step involves selecting the variables affecting the outcome from the set of covariates identified in the first step. de Luna and Waernbaum use a polynomial logistic regression in the first stage and a second order polynomial regression in the second stage. In both stages, the covariate selection is based on a forward stepwise procedure and Akaike's information criteria. We implement a slightly simplified approach, using a probit model in the first step and OLS in the second. In both stages, we use forward stepwise selection and a minimum significance level of 10 percent for adding variables to the model. Set of dummy variables indicating more than two categories are added on the basis of their joint statistical significance. We start with a model containing only quadratics in grade point average and study program in upper secondary school.

Table 6 presents the estimated effects using single nearest neighbor matching with replacement and different specifications of the propensity scores according to the procedure described above. The number of variables used in each specification is reported in brackets (relative to the number of variables in the corresponding full specification). Three main findings emerge from the table. First, the applied covariate selection approach produces propensity score specifications with significantly fewer conditioning variables. This is especially true for the sub-samples for specific college fields/majors. Among the covariates which are typically selected, we find (apart from the pre-specified upper secondary school grade point average and study program) pre-treatment region of residence, educational level of the father, college field/major and degree level.

The second finding is that the results are remarkably robust with regard to the different specifications of the propensity scores (compare with Table 3). The estimated effects continue to be very small and statistically insignificant at conventional levels.

The third finding is that the balancing indicators typically show somewhat better overall covariate balancing as compared to the achieved balancing for the full propensity score specifications. This is not surprising, given that the number of conditioning variables is substantially smaller.

Table 6. Estimated effects of college choice on annual earnings with different specifications of the propensity score

	All majors		Law and social sciences		Technology	
	Men	Women	Men	Women	Men	Women
Nearest neighbor(1);	0.008	0.004	−0.001	−0.007	−0.000	−0.018
step 1	(0.0114)	(0.0108)	(0.0276)	(0.0232)	(0.0127)	(0.0315)
	[25/34]	[28/34]	[22/28]	[21/28]	[17/28]	[16/28]
Balancing indicators						
Mean bias before	26.8	27.6	29.8	29.2	34.0	36.6
Mean bias after	2.4	3.8	6.0	4.7	3.3	4.5
Pseudo R^2 before	0.259	0.237	0.287	0.263	0.265	0.282
Pseudo R^2 after	0.004	0.009	0.014	0.012	0.006	0.010
Nearest neighbor(1);	0.008	−0.001	0.025	−0.020	−0.006	−0.031
step 2	(0.0109)	(0.0111)	(0.0192)	(0.0198)	(0.0128)	(0.0316)
	[25/34]	[20/34]	[11/28]	[16/28]	[12/28]	[12/28]
Balancing indicators						
Mean bias before	26.8	28.4	30.8	29.7	34.4	37.1
Mean bias after	2.4	3.3	3.9	4.4	2.5	4.9
Pseudo R^2 before	0.259	0.224	0.174	0.238	0.260	0.277
Pseudo R^2 after	0.004	0.006	0.005	0.008	0.003	0.008
Number of observations	21,473	23,334	6,192	7,066	9,711	2,713

Notes: The outcome variable is the log of total annual earnings from employment and self-employment in 2003. The propensity scores are estimated using a probit model. The specification of the propensity scores is based on a two-step covariate selection procedure. We refer to the text for details on this. Bootstrap standard errors based on 250 replications appear in parentheses below the estimates and the number of variables used in each specification is reported in brackets (relative to the number of variables in the corresponding full specification).

8. Summary and concluding remarks

The purpose of this paper has been to estimate the causal effect on earnings of graduating from *old universities* rather than *new universities/university colleges*. We have used several different propensity score matching methods and linear regression. The overall conclusion from the analysis is that we cannot find any significant differences in earnings between graduates from the two groups of colleges. This holds for male and female sub-samples covering all majors, as well as male and female sub-samples covering two broad fields of education. We find that the results are robust with regard to different methods of propensity score matching and regression adjustment. The results furthermore indicate little sensitivity with regard to the empirical support in the data and alternative specifications of the propensity scores. In effect, this means that the unconditional earnings premium of about 8−15 percent (depending on the sub-sample) of graduating from old universities, disappears when we compare comparable individuals using different types of propensity score matching methods and linear regression. These findings confirm the results in Eliasson (2006).

Unlike the majority of the papers in the international literature, this study has not focused on the labor market effects of college quality as such. Yet, we know from the introduction that there are major differences between the two groups of colleges in terms of factors presumably related to college quality. The percentage of faculty with doctoral degrees is almost twice as high at old universities and, as the name implies, this group of universities is also characterized by long standing academic traditions. So why do these favorable conditions at the old universities not translate into higher earnings of their college graduates? One harsh answer to this question is that the old universities simply do not produce education of higher quality than the new universities/university colleges, despite their superior endowment. A perhaps more plausible explanation is that the link between educational quality and earnings is especially week in the highly organized Swedish labor market, known for its narrow wage distribution. However, during the last 10 to 15 years the Swedish bargaining system has undergone dramatic changes, including a strong trend towards individual wage setting and rapidly increasing wage differentials among white-collar workers (Lundborg, 2005). This suggests that traditional market forces indeed have influenced wage setting during the period in focus here. To some extent, this challenges the argument of a particularly week relationship between earnings and educational quality in the Swedish labor market.

Compared to the many studies on the economic returns to education in terms of years of schooling completed or level of education attained, the number of papers focusing on the labor market effects of college choice is still very limited. To increase the knowledge about the relationship between college choice and labor market outcomes, more studies, based on different data sources and identifications strategies, are needed. One important topic for future research is to introduce more detailed measures of college quality into the analysis. Another useful research issue is to evaluate the effect of college choice on other outcome variables than earnings. Employment opportunities and unemployment risks are examples of interesting candidates. All together, such developments will contribute to our understanding about the relationship between college choice and labor market outcomes.

References

Antelius, J. and Björklund, A. (2000), How Reliable are Register Data for Studies of the Return on Schooling? An Examination of Swedish Data, *Scandinavian Journal of Education Research 44* (4), 341−355.

Augurzky, B. and Schmidt, C. (2001), The Propensity Score: A Means to An End, Discussion Paper no. 271, Institute for the Study of Labor, Bonn.

Behrman, J., Rosenzweig, M. and Taubman, P. (1996), College Choice and Wages: Estimates Using Data on Female Twins, *Review of Economics and Statistics 78* (4), 672−685.

Berg Dale, S. and Krueger, A. (2002), Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables, *Quarterly Journal of Economics 117* (4), 1491−1527.

Black, D. and Smith, J. (2004), How Robust is the Evidence on the Effects of College Quality? Evidence from Matching, *Journal of Econometrics 121* (1−2), 99−124.

Black, D. and Smith, J. (2006), Estimating the Returns to College Quality with Multiple Proxies for Quality, *Journal of Labor Economics*, forthcoming.

Black, D., Daniel, K. and Smith, J. (1995), College Characteristics and the Wages of Young Women, unpublished manuscript, University of Maryland.

Black, D., Daniel, K. and Smith, J. (1997), College Quality and the Wages of Young Men, unpublished manuscript, University of Maryland.

Black, D., Daniel, K. and Smith, J. (2005), College Quality and Wages in the United States, *German Economic Review 6* (3), 415−443.

Brewer, D. and Ehrenberg, R. (1996), Does it Pay to Attend an Elite Private College? Evidence from the Senior High School Class of 1980, *Research in Labor Economics 15*, 239−271.

Brewer, D., Eide, E. and Ehrenberg, R. (1999), Does it Pay to Attend an Elite Private College? Cross-Cohort Evidence on the Effects of College Type on Earnings, *Journal of Human Resources 34* (1), 104−123.

Caliendo, M. and Kopeinig, S. (2006), Some Practical Guidance for the Implementation of Propensity Score Matching, *Journal of Economic Surveys*, forthcoming.

Card, D. (1999), The Causal Effect of Education on Earnings, in O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics*, Vol. 3A, Elsevier, Amsterdam.

Chevalier, A. and Conlon, G. (2003), Does it Pay to Attend a Prestigious University?, CEE Discussion Papers no. 0033, London School of Economics and Political Science.

Datcher Loury, L. and Garman, D. (1995), College Selectivity and Earnings, *Journal of Labor Economics 13* (2), 289−308.

Dehejia, R. and Wahba, S. (1999), Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs, *Journal of the American Statistical Association 94* (448), 1053−1062.

Dehejia, R. and Wahba, S. (2002), Propensity Score-Matching Methods for Nonexperimental Causal Studies, *Review of Economics and Statistics 84* (1), 151−161.

de Luna, X. and Waernbaum, I. (2005), Covariate Selection for Non-Parametric Estimation of Treatment Effects, Working Paper no. 2005:4, Institute for Labour Market Policy Evaluation, Uppsala.

Eliasson, K. (2006), The Role of Ability in Estimating the Returns to College Choice: New Swedish Evidence, Umeå Economic Studies no. 691, Umeå University.

Gartell, M. and Regnér, H. (2002), Arbetsmarknaden för högskoleutbildade. Inkomstutveckling och geografisk rörlighet under 1990-talet (The Labor Market for College Educated. Development of Earnings and Geographical Mobility During the 1990s), Swedish Confederation of Professional Associations, Stockholm.

Gartell, M. and Regnér, H. (2005), Sambandet mellan val av högskola och inkomster efter examen för kvinnor och män (The Relation Between College Choice and Subsequent Earnings for Women and Men), Rapport 2005:12, Institute for Labour Market Policy Evaluation, Uppsala.

Heckman, J. and Robb, R. (1985), Alternative Methods for Evaluating the Impact of Interventions, in J. Heckman and B. Singer (eds.), *Longitudinal Analysis of Labor Market Data*, Cambridge University Press, New York.

Heckman, J., Ichimura, H. and Todd, P. (1997), Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme, *Review of Economic Studies 64* (4), 605−654.

Heckman, J., Ichimura, H. and Todd, P. (1998a), Matching as an Econometric Evaluation Estimator, *Review of Economic Studies 65* (2), 261−294.

Heckman, J., Ichimura, H., Smith, J. and Todd, P. (1998b), Characterizing Selection Bias Using Experimental Data, *Econometrica 66* (5), 1017−1098.

Heckman, J., LaLonde, R. and Smith, J. (1999), The Economics and Econometrics of Active Labor Market Programs, in O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics*, Vol. 3A, Elsevier, Amsterdam.

Imbens, G. (2004), Nonparametric Estimation of Average Treatment Effects under Exogenity: A Review, *Review of Economics and Statistics 86* (1), 4−29.

King, G. and Zeng, L. (2006), The Dangers of Extreme Counterfactuals, *Political Analysis 14* (2), 131−159.

Leuven, E. and Sianesi, B. (2003), PSMATCH2: Stata Module to Perform Full Mahalanobis and Propensity Score Matching, Common Support Graphing, and Covariate Imbalance Testing, http://ideas.repec.org/c/boc/bocode/s432001.html.

Lindahl, L. and Regnér, H. (2005), College Choice and Subsequent Earnings: Results Using Swedish Sibling Data, *Scandinavian Journal of Economics 107* (3), 437−457.

Lundborg, P. (2005), Individual Wage Setting, Efficiency Wages and Productivity in Sweden, Working Paper Series no. 205, Trade Union Institute for Economic Research, Stockholm.

Lundin, M. (2006), Effects of College Choice on Income: Estimation and Sensitivity Analysis, Licentiate Thesis no. 35, Department of Statistics, Umeå University.

Monks, J. (2000), The Returns to Individual and College Characteristics. Evidence from the National Longitudinal Survey of Youth, *Economics of Education Review 19* (3), 279−289.

Rosenbaum, P. (1984), The Consequences of Adjustment for a Concomitant Variable That Has Been Affected by the Treatment, *Journal of the Royal Statistical Society A 147* (5), 656−666.

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

Rosenbaum, P. and Rubin, D. (1985), Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score, *The American Statistician 39* (1), 33−38.

Rubin, D.B. and Thomas, N. (1996), Matching Using Estimated Propensity Scores: Relating Theory to Practice, *Biometrics 52* (1), 249−264.

Sianesi, B. (2004), An Evaluation of the Swedish System of Active Labor Market Programs in the 1990s, *The Review of Economics and Statistics 86* (1), 133−155.

Smith, J. (1997), Matching with Multiple Controls to Estimate Treatment Effects in Observational Studies, *Sociological Methodology 27* (1), 325−353.

Smith, J. (2000), A Critical Survey of Empirical Methods for Evaluating Active Labor Market Policies, *Zeitschrift für Volkswirtschaft und Statistik 136* (3), 1−22.

Smith, J. and Todd, P. (2005a), Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators?, *Journal of Econometrics 125* (1−2), 305−353.

Smith, J. and Todd, P. (2005b), Rejoinder, *Journal of Econometrics 125* (1−2), 365−375.

Willis, R. (1986), Wage Determinants: A Survey and Reinterpretation of Human Capital Earnings Functions, in O. Ashenfelter and R. Layard (eds.), *Handbook of Labor Economics*, Vol. 1, Elsevier, Amsterdam.

Zhang, L. (2006), Do Measures of College Quality Matter? The Effects of College Quality on Graduates' Earnings, *The Review of Higher Education*, forthcoming.

Öckert, B. (2001), Effects of Higher Education and the Role of Admission Selection, Ph.D. Thesis no. 52, Swedish Institute for Social Research, Stockholm University.

Öckert, B. and Regnér, H. (2000), Högre utbildning i Sverige. En problemorienterad diskussion om utbildningssatsningar (Higher Education in Sweden: A Discussion on Investments in Education), Swedish Institute for Social Research, Stockholm University.

Appendix A. College grouping (by year of establishment)

Old Universities	New universities/university colleges	
Uppsala University (1477)	Stockholm Institute of Education (1956)	Mälardalen University College (1977)
Lund University (1666)	Stockholm University College of Physical Education and Sports (1966)	University College of Borås (1977)
Karolinska Institutet (1810)	Umeå University (1965)	University College of Gävle (1977)
KTH – Royal Institute of Technology (1826)	Luleå University of Technology (1971)	University College of Kalmar (1977)
Chalmers University of Technology (1829)	Linköping University (1975)	Växjö University (1977)
Stockholm University (1878)	Dalarna University College (1977)	Örebro University (1977)
Göteborg University (1891)	Jönköping University College (1977)	Halmstad University College (1983)
Stockholm School of Economics (1909)	Karlstad University (1977)	University College of Skövde (1983)
	Kristianstad University College (1977)	Blekinge Institute of Techonology (1989)
	Mid Sweden University College (1977)	University College of Trollhättan/Uddevalla (1990)
		Malmö University College (1998)

Notes: The division is based on the official status of the colleges in 1999. Year of establishment in parentheses. In some cases, the colleges began providing limited education a few years earlier than reported in the table. Karlstad, Växjö and Örebro were originally established as university colleges, but received official status as universities in 1999.