Società italiana di economia pubblica

XIX CONFERENZA

ECONOMIA DEL CAPITALE UMANO
Istituzioni, incentivi e valutazioni

Pavia, Aule storiche Università, 13 - 14 settembre 2007

EARLY LABOUR MARKET RETURNS TO COLLEGE SUBJECT

PAOLO BUONANNO AND DARIO POZZOLI

pubblicazione internet realizzata con contributo della COMPAGNIA di San Paolo

Early Labour Market Returns to College Subject

Paolo Buonanno* and Dario Pozzoli†

Department of Economics, University of Bergamo

Abstract

This paper aims at estimating early labour market returns (i.e. participation probability, employment probability and log hourly earnings) of Italian university graduates across college subjects. We devote great attention to endogenous selection issues using alternative methods to control for potential self-selection associated with the choice of the degree subject in order to unravel the causal link between college major and subsequent outcomes in the labour market. We use both a propensity score matching-average treatment on the treated method and the polychotomous selectivity model introduced by Lee (1983) to investigate the existence of unobserved heterogeneity. Our results suggest that "quantitative" fields (i.e. Sciences, Engineering and Economics) increase not only participation to the labour market and employment probability but also early earnings, conditional on employment.

* *Corresponding Author*. Address: Dipartimento di Scienze Economiche, Università degli Studi di Bergamo, Via dei Caniana 2, 24127 Bergamo, Italy. *Email*: paolo.buonanno@unibg.it. Usual disclaimers apply.

† Address: Dipartimento di Scienze Economiche, Università degli Studi di Bergamo, Via dei Caniana 2, 24127 Bergamo, Italy. *Email*: dario.pozzoli@unibg.it.

1. Introduction

"…While sending your child to Harvard appears to be a good investment, sending him to your local state university to major in Engineering, to take lots of math, and preferably to attain a high GPA, is even a better private investment" (James et al. 1989, p. 252).

Over the last 40 years a large body of research has focussed on the economic returns to higher education. However, the vast majority of these studies estimate the average return to education without controlling for the degree subject.

A number of previous works both for the US and for the UK (Daymont and Andrisani 1984; Berger 1988; James et al. 1989; Grogger and Eide 1995; Loury and Garman 1995; Loury 1997; Blundell et al. 2000) document the large differences in earnings across fields of study, but none of these papers model the choice of college subject taking into account the issue of self-selection.

Recently, Arcidiacono (2004) developing a dynamic model of college and major choice, that allows to control for selection, shows that large earnings differences exist across majors. Similarly, Bratti and Mancini (2003), focus on the early occupational earnings of young UK graduates by adopting different methodological approaches, and estimate the wage *premia* across different college subjects.

The aim of this paper is to investigate the differences in early labour market outcomes (i.e. participation probability, employment probability and log hourly earnings three years after graduation) between Italian university graduates across college major using alternative methods to control for potential self-selection associated with the choice of the degree subject.

We consider a multiple treatment model, which distinguishes the impact of the different university groups, thus allowing the attainment of different educational qualifications to have separate effects.

We devote great attention to endogenous selection issues in order to unravel the casual link between field of study and subsequent outcomes in the labour market, using both a propensity score matching method (Rosenbaum and Rubin 1983) and the polychotomous selectivity model (Lee 1983) to account for the existence of unobserved heterogeneity.

For our empirical analysis we use two waves (2001 and 2004) of the Graduates' Employment Survey (GES) conducted by the Italian National Statistical Institute (ISTAT) three years after graduation.

The economic returns for Italian university graduates have been extensively investigated in Italian empirical works (Biggeri et al. 2000; Boero et al. 2004; Makovec 2005; Cappellari and Brunello 2007). However, none of the previous studies have explicitly modelled the choice of college subject taking into account the issue of self-selection.[1]

Our results suggest that "quantitative" fields (i.e. Sciences, Engineering and Economics) increase not only participation to the labour market and employment probability but also early earnings, conditional on employment. Graduates in Humanities and Social Sciences are always the most disadvantaged in terms of

[1] Ballarino and Bratti (2006) represents a notable exception, even if they focus on the effect of different fields of study in the university to work transition.

employment probability and have generally a negative earning *premium* with respect to graduates from the other subjects.

The rest of the paper proceeds as follows. Section 2 describes the empirical methodology. Data as well as model specification are presented in Section 3. Section 4 provides the empirical results and estimates the *premia* for different college subjects. Section 5 concludes.

2. Empirical Methodology

In this section, we present the econometric methodology used to estimate the labour market returns to university degree across college subjects. We consider the effect of a multiple treatment, namely college major, on participation probability, employment probability and log hourly wages.

We estimates labour market *premia* comparing labour market outcomes for individuals who graduated in one subject with "matched" individuals who graduated in a different major. This approach consider the college major as the treatment that the individual receives and aims at assessing the effect of this treatment on the outcome variables.

The general matching method is a non-parametric approach to the problem of identifying the treatment impact on outcomes. To recover the average treatment effect on the treated (ATT), the matching method tries to mimic *ex-post* an experiment by choosing a comparison group between the non-treated such that the selected group is as similar as possible to the treatment group in terms of their observable characteristics. Under the matching assumption, all the outcome-relevant differences between treated and non-treated individuals are captured in their observable attributes, the only remaining difference between the two groups being their treatment status. The central issue in the matching method is the choice of the appropriate matching variables.

Following Lechner (2001), the multiple evaluation problem can be presented as follows.[2]

Consider participation in (M+1) mutually exclusive treatments, denoted by an assignment indicator D $\in$ (1,…,M). In our case we assume that an individual can choose among five different alternatives D $\in$ (1,…,5), which are: (1) Sciences, (2) Engineering, (3) Economics, (4) Social Sciences and (5) Humanities.[3] X denote the set of variables unaffected by treatments, while the outcome variables are denoted by $(Y^0,\ldots,Y^M)$. Each individuals receives exactly one of the treatments, therefore for any participants, only one component of $(Y^0,\ldots,Y^M)$ can be observed in the data. The remaining M outcomes represent counterfactuals. The number of observations in the population is N, such that $N = \sum_{m=0}^{M} N^m$, where N^m is the number of participants in treatment m. The focus is on a pair-wise comparison of the effects of treatment m and l, for all combinations of m, l $\in$ (0,1,…,M), m≠l. More formally, the outcome of interest in this study is presented by the following equation:

$$\theta_0^{ml} = E(Y^m - Y^l \mid D = m) = E(Y^m \mid D = m) - E(Y^l \mid D = m) \qquad (2.1)$$

[2] For a complete description of the methodology used in our paper we refer to Lechner (2001).

[3] We present the composition of each university group in the following section.

θ_0^{ml} in equation (2.1) denotes the expected average treatment effect of treatment m relative to treatment l for participants in treatment m (sample size N^m).

The evaluation problem is a problem of missing data: one cannot observe the counterfactual $E(Y^l|D=m)$ for m≠l, since it is impossible to observe the same individual in several states at the same time. Thus, the true causal effect of a treatment m relative to treatment l can never be identified. However, the average causal effect described by equation (2.1) can be identified under the conditional independence assumption (CIA).[4]

Moreover, for the average treatment effect to be identified, the probability of treatment m has to be strictly between zero and one,[5] i.e.

$$0<P^m(X)<1, \text{ where } P^m(x)=E[P(D=m \mid X=x)],\ \forall\ m=0,1,\ldots,M \quad (2.2)$$

which prevents X from being perfect predictors of treatment status, guaranteeing that all treated individuals have a counterpart in the non-treated population for the set of X values over which we seek to make a comparison.

As discussed in Lechner (2001), the balancing score property, suggested by Rosenbaum and Rubin (1983) for the binary case, holds for multiple case as well:

$$(Y^0,\ldots,Y^M) \perp D \mid X=x,\ \forall\ x \in X \text{ if } (Y^0,\ldots,Y^M) \perp D \mid b(X)=b(x),\ \forall\ x \in X \quad (2.3)$$

The main advantage of the balancing score property is the decrease in dimensionality: instead of conditioning on all the observable covariates, it is sufficient to condition on some function of the covariates. In the case of multiple treatments, a potential and quite intuitive balancing score is the M-dimensional vector of propensity scores $[P^0(x), P^1(x),\ldots,P^M(x)]$.

To identify and estimate θ_0^{ml}, first of all we identify and estimate $E(Y^m|D=m)$ by the sample mean. The conditional independence assumption implies that the latter part of equation (2.3), $E(Y^l|D=m)$, is identified in large enough samples as:

$$E[\ E(Y^l \mid b(X), D=m) \mid D=m] = E(Y^l \mid D=m) \quad (2.4)$$

To estimate (4.4), Imbens (2000) and Lechner (2001) show that instead of M-dimensional balancing score the dimension of the conditioning set can be reduced to $[P^m(x), P^l(x)]$. Thus,

$$E(Y^l \mid D=m) = E[\ E(Y^l \mid P^m(X), P^l(X), D=l) \mid D=m] \quad (2.5)$$

We decide to model this choice using a multinomial logit model. The probability that an individual I, with the set of characteristics X_i choose the subject *m* is give by the following expression:

[4] CIA states that all differences affecting the selection between the groups of participants in treatment m and treatment l are captured by observable characteristics X. In the multiple case as presented in this paper, the CIA is formalised as follows $(Y^0,\ldots\ldots,Y^M) \perp D \mid X=x,\ \forall\ x \in X$.

[5] This is also known as the common or overlap condition. Depending on the sample in use, this can be quite a strong requirement and the estimated treatment effect has then to be redefined as the mean treatment effect for those treated falling within the common support.

$$\Pr(D_i = m) = \frac{\exp(X_i \eta_m)}{\sum_{l=0}^{4} \exp(X_i \eta_l)} \qquad (2.6)$$

where X_i include pre-treatment variables: family background characteristics (parents education, father occupation), academic performance at high school, high school type, age, region of residence, survey year, compulsory military service before university and change in living arrangements.

3. Data description and model specification

Our data originate from the 2001 and 2004 waves of the Graduates' Employment Survey (GES) conducted by the Italian National Statistical Institute. The sample, consisting of approximately 5 percent of the population of Italian university graduates, is representative of students who got their college degree in 1998 and 2001.[6] The surveys collect a wide range of information on academic curriculum, post-graduate labour market experiences, personal characteristics and family background for a representative sample of 46,850 Italian university graduates. The data allows in particular tracking the whole educational history of each individual, and provides a full description of academic and labour market performance during the three years after their graduation.[7]

The list and the definition of the variables, together with summary statistics, are presented in Table 1. The university groups have been classified into 5 main categories: Sciences (Chemistry, Physics, Geology, Biology, Pharmacy, IT and Mathematics); Engineering (including Architecture); Economics (including Statistics and Business); Social Sciences (Sociology, Political Sciences, and Law) and Humanities (Philosophy, Literature, Languages, Education, Psychology).[8]

As far as model specification is concerned, we present the set of socio-demographic and education variables use in our analysis. It is important to note that all the matching variables included in the selection equation (2.6) are variables that are unaffected by college choice enrolment, because fixed over time or measured before enrolment at university.

In our empirical analysis we exploit the following information contained in the surveys. Individual characteristics include sex, age, region of residence,[9] if the

[6] Response rate in both surveys is around 60%.

[7] For the present analysis, the sample of 46,850 records has been reduced by eliminating those: i) who were employed and started their job while at university, since their post-graduation experiences might not be comparable with those of the rest of the sample; ii) for whom information on earnings is missing; iii) who graduated from Medicine and physical training. The resulting sample size is nearly 35,000 high school leavers, of whom nearly 27,000 participate to the labour market and 21,504 are full-time employed three years after graduation.

[8] Due to the complexity of the model and the number of parameters to be estimated we were not able to consider a finer definition of college majors. A similar level of aggregation is used in most of the studies reviewed in the previous section (Berger 1988; Rochat and Demeulemeester 2001; Bratti and Mancini 2003; Arcidiacono 2004).

[9] Makovec (2005) and Brunello and Cappellari (2007) document that that the percentage of individuals who do not move to attend university is close to three quarters of the population of graduates. For

individual attended university in a city different from the one where she lived (commuter within the same region) and whether the individual did the compulsory military service before college. Indicators of past educational choices and performance are the type of high school degree obtained and the high school final mark. Family background variables include both parents' education[10] (with a breakdown in university, high school and primary school) and fathers' occupation (with a breakdown in entrepreneur, professional, manager, high skilled and low skilled white collar, blue collar, other independent and no qualifications).

Table 2 presents graduates distribution according to college major, and show that the (weighted) sample provides a very good representation of the population. Graduates from Sciences and Engineering represent 15% and 18% of the whole sample respectively, while graduates from Humanities and Social Sciences constitute approximately 23%. Finally, Economics graduates represent 20%.

Table 3 reports the distribution of graduates by high school final mark (an indicator of student's performance and ability). Nearly 30% and 25% of students with high grades (above 56/60) in high school got respectively a degree in Engineering and in Economics; while those students who performed low grades (below 40) instead are less likely to graduate from this field of study (only 9%). Table 4 presents the distribution of graduates by high school type: 50% and 37% of Economics and Engineering graduates got scientific general high school degree; while most of the students from Humanities and Social Sciences come from humanities general high school (30%).[11]

We conduct our empirical analysis for three different labour market outcome variables: log net hourly earnings for full-time employed, participation to the labour market and full-time employment. [12]

Wages are available for approximately 20,600 individuals and their distribution is presented in Figure 1.[13] From Table 5, that shows the average wage by university groups, it clearly emerges that the average outcome measure is highest for graduates in Engineering and Economics. Table 5 also reports the distribution of the participation and employment rates by college major. On the one hand, graduates from Engineering and Economics decide to participate to the labour market (about 90% and 81% respectively), of whom nearly 80% are full-time employed. On the other hand it is interesting to see that 74% of graduates from Humanities participate to the labour market after graduation and only 60% of them have found a job three years after graduation.

---

instance, 71% of our sample did not move to attend college. Hence we can consider the region of university as a good proxy for the region of residence before enrolling at university.

[10] We decide to model in this way the parental education to capture the main interaction effects due to the assortative mating behaviour (Behrman and Rosenzweig 2002).

[11] The classification of high schools used in the analysis is the following: scientific general high school (liceo scientifico); humanities general high school (liceo classico); vocational high school (istituti tecnici e professionali); other high schools (istituti magistrali, liceo artistico, istituto d'arte; altra maturità).

[12] It is important to note that in the 2001 survey the earnings are available only in a interval-censored form. We obtain the continuous variable through the interval regression model (see Stewart 1983; Bryson 2002).

[13] We dropped from the original sample the extreme observations of the monthly earnings and of the hours worked per week (those lower than 1[th] percentile of the earnings/hours distributions and those higher than 99[th] percentile). The log hourly earnings are available for 20,600 individuals.

4. Empirical Results

This section provides the results from estimating the labour market *premia* by college majors. In the first subsection we discuss balancing score property, the average treatment effect on the treated is shown in subsection 4.2, while subsection 4.3 presents the robustness of our findings to different methodologies.

4.1 Matching

Table 5.1-5.3 report covariate balancing indicators for pair-wise matching.

In the pair-wise matching, each individual in the treated sub-sample m is matched with a comparison in the sub-sample l, and the criteria for finding the nearest possible match is to minimise the Mahalanobis distance of $[P^m(X), P^l(X)]$ between the two units. The covariance matrix for the estimates of the average effects, suggested and presented by Lechner (2002), pays regard to the risk of over-using some of the comparison unit: the more times each comparison is used, the larger the standard error of the estimated average effect. We opt to estimate the standard errors with the bootstrapping methodology.[14]

Furthermore, covariates in the matched samples ought to be balanced according to the condition $X \perp D \mid b(X)$. Following Lechner (2001), the match quality is judged by the mean absolute standardized biases of covariates. The results reported in Tables 5.1-5.3 show that, in general, a satisfactory matching is achieved for the reported model specifications and for the different sub-samples, and thus the condition $X \perp D \mid b(X)$ is fulfilled.

4.2 Average treatment on the treated effects.

In this section, we firstly estimate, on the entire sample of graduates, the probability of participation to the labour market by field of study. Then, for those individuals who decide to participate to the labour market, we estimate the probability of employment. Finally, after dropping out from the sample those individuals who are unemployed, we estimate the average matching treatment on the treated effects on earnings conditional on employment.

We seek to ensure the quality of matches by setting different tolerance levels when comparing propensity scores (i.e. we impose two different calipers: 0.01 and 0.001). Imposing a caliper work in the same direction as allowing for replacement: bad matches are avoided and hence the matching quality raises. Furthermore, by setting different calipers we can check the robustness of our results to different common support definitions.

Each estimated effect is reported in relative terms expressed in percentage in Tables 6-8. Our findings show higher labour market *premia* both for Economics and Engineering graduates (Table 8). The wage *premia* of Economics and Engineering relative to Humanities are respectively between 6% and 8%, while are significantly higher relative to Humanities (respectively 12% and 15%). Economics and Engineering increase not only earnings but also participation to the labour market (Table 6) and employment probability (Table 7). For instance, graduates from Economics and

[14] Standard errors are bootstrapped using 500 replications (see Black and Smith 2003; Sianesi 2004).

Engineering present an employment rate that is respectively 25% and 10% higher relative to Humanities and Social Sciences graduates and 10%. While graduates in Sciences have higher labour market outcomes than Humanities and Social Sciences, but show lower employment probability and participation rate with respect to graduates from Economics and Engineering.

Overall, "quantitative" fields (Engineering, Economics and Sciences) increase not only participation to the labour market and employment probability but also early earnings, conditional on employment

4.3 Sensitivity analysis for the average treatment effects on earnings.

This section reports the robustness checks of the results for the earnings *premia* by degree subject presented in the previous section.

Firstly, the sensitivity of the results to the methodology used to estimate the average treatment effect of fields of study on hourly earnings is investigated. Even though matching is a relatively flexible and above all intuitive method to compare the effects of various treatments and to explore the extent of heterogeneity in the treatment effect among the individuals, it has some drawbacks. On the one hand, the assumption of conditional independence is not only very strong but also impossible to test. On the other hand, even though we do not need to specify the outcome model, we need to be careful about the specification of the discrete choice model, the criterion of matching, and the definition of common support. Hence, in this section we introduce a different approach to determine the average treatment effect on earnings and relate it to the propensity score matching method and to the results presented in the previous section. In particular, we utilize the polychotomous selectivity model introduced by Lee (1983) to investigate the existence of unobserved heterogeneity.

Secondly, we investigate the sensitivity of the results to the specification and estimation of the propensities.

4.3.1 Polychotomous selectivity model

The model presented by Lee (1983) is designed for dealing with selectivity bias in the polychotomous case when the dependent variable is continuous. The idea of this approach is largely the same as in the approach introduced by Dubin and McFadden (1984), which in turn is a multinomial generalisation of Heckman's two stage method.[15] Like all these selectivity models, the Lee's model is designed to adjust for both observed and unobserved selection bias. Thus, it does not require the conditional independence assumption to be valid.[16] Consider the following model:

$$y_1 = x\beta_1 + u_1$$
$$y_m^* = z\gamma_m + \eta_m, \text{ m=0,…..,M (4.1)}$$

[15] The main shortcoming of the Lee approach compared to the one presented by Dubin and McFadden (1984) is that it contains relatively restrictive assumptions on the covariance between the error term ε and μ.

[16] However, it rests on other strong assumptions, among them linearity in the outcome variable and joint normality in the error terms.

where the disturbance u_1 is not parametrically specified and verifies $E(u_1 \mid x, z) = 0$ and $V(u_1 \mid x,z) = \sigma^2$; y_m is a categorical variable that describes the choice of an economic agent among M alternatives based on "utilities" y_m^*.[17] The vector z represents the maximum set of explanatory variables for all alternatives and the vector x contains all determinants of the variable of interest. It is assumed that the model is non-parametrically identified from exclusion of some of the variables in z from the variables in x. Hence this approach attempts to control for selection on unobservables by exploiting some exogenous variation in schooling through some excluded instruments. Our data set contains the information on the number of siblings that has been often considered a potential instrument in the related literature (see Haveman and Wolfe, 1995). Hence this variable may determine assignment to college major but, conditional on the x_s, could be excluded from the earnings equation.

Without loss of generality, the outcome variable y_1 is observed if and only if category 1 is chosen, which happens when:

$$y_1^* > \max_{\neq 1}(y_j^*) \quad (4.2)$$

Define

$$\varepsilon_1 = \max_{m\neq 1}(y_m^* - y_1^*) = \max_{m\neq 1}(z\gamma_m^* + \eta_m - z\gamma_1 - \eta_1) \quad (4.3)$$

Under definition (4.3), condition (4.2) is equivalent to $\varepsilon_1<0$. Assume that the (η_m)'s are independent and identically Gumbel distributed (the so-called IIA hypothesis). As shown by McFadden (1973), this specification leads to the multinomial logit model. Based on this assumption, consistent maximum likelihood estimates of (γ_m) 's can be easily obtained. The problem is to estimate the parameter vector β_1 while taking into account that the disturbance term u_1 may not be independent of all (η_j)'s. This would introduce some correlation between the explanatory variables and the disturbance term in the outcome equation model (4.1). Because of this, least squares estimates of β_1 would not be consistent. Lee (1983) proposed a generalization of the two-step selection bias correction method introduced by Heckman (1979) that allows for any parameterized error distribution. His method extends to the case where selectivity is modelled as a multinomial logit. This approach is simple and requires the estimation of only one parameter in the correction term. This is however achieved at the cost of fairly restrictive assumptions (Lee 1983). [18]

---

[17] The choice alternatives are: sciences (m=1), engineering (m=2), economics (m=3), social sciences (m=4) and humanities (m=5).

[18] Call $F_{\varepsilon 1}(.|\Gamma)$ the cumulative distribution function of ε_1. The cumulative $J_{\varepsilon 1}(.|\Gamma)$, defined by the following transform: $J_{\varepsilon 1}(.\mid\Gamma) = \Phi^{-1}(F_{\varepsilon 1}(.\mid\Gamma))$, where Φ is the standard normal cumulative, has a standard normal distribution. Assume that u_1 and $J_{\varepsilon 1}(.\mid\Gamma)$ are linearly related with correlation ρ_1 (this holds in particular if they are bivariate normal). Then, the expected value of the disturbance term u_1, conditional on category 1 being chosen, is given by: $E(u_1 \mid \varepsilon_1 < 0, \Gamma) = -\sigma\rho_1 \dfrac{\phi(J_{\varepsilon 1}(0\mid\Gamma)}{F_{\varepsilon 1}(0\mid\Gamma)}$.

Under these assumptions, a consistent estimator of β_1 is obtained by estimating least squares of the following equation:

$$y_1 = x_1\beta_1 - \sigma\rho_1 \frac{\phi(J_{\varepsilon 1}(0 \mid \Gamma)}{F_{\varepsilon 1}(0 \mid \Gamma)} + w_1 \quad (4.4)$$

Two-step estimation of (4.4) is thus implemented by first estimating the (γ_j)'s in order to obtain $\frac{\phi(J_{\varepsilon 1}(0 \mid \Gamma))}{F_{\varepsilon 1}(0 \mid \Gamma)}$ and then by including that variable in equation (4.4) to estimate consistently β_1 and $\sigma\rho_1$ by least squares.

The results in Table 10 show that including the selection adjustment terms in the equation for earnings produces much higher estimates in absolute value of the ATTs compared to the matching ones.[19] This is not surprising since identification is based on a different assumption, i.e. the individuals are allowed to select into college major on the basis of their idiosyncratic gains. Moreover these differences are presumably explained by the parametric restrictions underlying the control function approach. However as in the matching framework, the results indicate that Humanities and Social Sciences graduates show a negative earning *premium* with respect to graduates from the "quantitative" fields. Our results are robust to accounting for unobserved heterogeneity through the polychotomous selectivity model. Furthermore, as suggested by our estimates, the parameters for selection adjustment terms are never statistically significant (Table 9). Hence, we find evidence suggesting that the matching approach with available set of Xs (i.e. observables) is not subject to selection bias.[20]

4.3.2 Binomial logit estimates

As we have argued in the main section, the multinomial logit estimates requires the Independence of Irrelavant Alternatives assumption to hold, i.e. the inclusion of new alternatives – or exclusion of some of the existing alternatives – does not alter the relative probability of a choice alternative to another. This assumption is convenient for estimation but not appealing from an economic or behavioural point of view.

In order to analyse the sensitivity of our multinomial logit results we utilize a matching procedure based on propensities obtained from binomial logit models. If the estimated coefficients in the binomial model are similar to the coefficients in the multinomial model the IIA assumption may be considered as valid. In general, the advantage of binomial logit compared to its multinomial counterpart is that it does not require validity of the IIA assumption. Nevertheless, Bryson et al. (2002) note that there are two shortcomings regarding this approach. First, as the number of options increases, the numer of models to be estimated increases disproportionately (for L options we need $0.5(L(L\text{-}1))$ models). Second, in each model only two options at a time are considered and consequently the choice is conditional on being in one of the two selected groups.

[19] Due to the presence of estimated coefficients in the creation of the counterfactual conditional means, we can easily surmise the correct standard deviations, through bootstrapping.

[20] It is interesting to note that under the structure imposed on the model, the estimated coefficients of the control functions are informative on the presence and direction of the selection process. Specifically, if an exclusion restriction can be found and the joint normality of the unobservables then the null hypothesis of no selection on the unobservables can be tested directly.

However, Lechner (2001) compares the performance of the multinomial probit approach and the series estimation and finds little difference in their relative performance. He suggests that the latter approach may be more robust since a mis-specification in one of the series will not compromise all others as would be the case in the multinomial probit model.

The vector of explanatory variables is identical to the one included in the multinomial model in the previous section. The binomial model is estimated for all five field choices and for the sample of employed graduates. The estimated coefficients are very similar to the coefficients in the multinomial model.[21] Thus, the IIA assumption appears to be fulfilled. The matching procedure is based on one-dimensional nearest-match criterion, i.e. each individual in sample m is matched with a comparison in sample l with the same, or nearest probability of treatment m, $P(T = m| X)$. As before, we impose different caliper specification to ensure the common support requirement. The matching quality measured by the absolute standardised bias of the samples is in general similar to the matching quality presented in the previous section. Table 11 presents results for the average treatment effect on the treated on log net hourly earnings.

The results do not qualitatively differ from our previous findings presented in Table 8. Matching on binomial propensities produces very similar effects to those estimated in the main section. The analysis presented in this section does not give reason to doubt the validity of the IIA assumption.

5. Concluding Remarks

The aim of this paper is to investigate the differences in early labour market outcomes (participation probability, employment probability and log hourly earnings three years after graduation) between Italian university graduates across college major. The analysis could be considered an advance in the literature because it does not limit itself to recognize that fields of study choice may be endogenous to the determination of early labour market outcomes, but attempts to correct directly for student self-selection into degree courses. To this end we employ both the propensity score matching technique which corrects for selectivity through observable characteristics and a simultaneous equation model of earnings determination and subject choice, which account for selectivity through unobservables (Lee 1983).

Using two waves (2001 and 2004) of the Graduates' Employment Survey conducted by the Italian National Statistical Institute (ISTAT) three years after graduation, we find that "quantitative" fields (i.e. Sciences, Engineering and Economics) increase not only participation to the labour market and employment probability but also early earnings, conditional on employment. Graduates in Humanities and Social Sciences are always the most disadvantaged in terms of early labour market outcomes. All these results may suggest that for those youths proceeding to the labour market after leaving university, quantitative fields offer better early labour market opportunities.

The last annual report of the Bank of Italy (2006) indicates that even if there exist huge differences in the employment returns of graduates by fields of study, the labour supply does not seem to adequate rapidly to the labour demand. Indeed, over the last 50

[21] Results are available upon request.

years the distribution of university graduates by fields of study has been almost stable with more than 60% of graduates from Humanities and Social Sciences and only one fourth from the "quantitative" subjects.

Our findings may be reconciled with the shortage in the supply of graduates in the quantitative field more than with skill biased technical change hypothesis, since both R&D expenditure are very low in Italy and graduates' employment opportunities have not changed during the two last decades (e.g. the structure of the Italian industry doesn't seem to favour the job market for high qualified technicians). This may be due to the fact that high school students decide to not enrol in the quantitative fields because they consider them a difficult and risky investment.

References

Arcidiacono, Peter. 2004. “Ability Sorting and the Returns to College Major.” Journal of Econometrics 121(1-2): 343-375.

Ballarino, Gabriele, and Massimiliano Bratti. 2006. “Fields of study and graduates’ occupational outcomes in Italy during the 90s. Who won and who lost?” Departmental Working Papers, Department of Economics, University of Milan.

Bank of Italy. 2007. “Relazione Annuale.” Roma: Banca d’Italia.

Behrman, Jere R., and Mark R. Rosenzweig. 2002. “Does Increasing Women’s Schooling Raise the Schooling of the Next Generation?” American Economic Review 92(1): 323-334.

Berger, Mark C. 1988. “Predicted future earnings and choice of college major.” Industrial and Labor Relations Review 41(3): 418-429.

Biggeri, Luigi, Matilde Bini and Leonardo Grilli. 2000. “The Transition from University to Work: A Multilevel Approach to the Analysis of the Time to Obtain the First Job.” Journal of the Royal Statistical Society Series A 164(2): 293-305.

Black, Dan A., and Jeffrey A. Smith, 2004. “How robust is the evidence on the effects of college quality? Evidence from matching.” Journal of Econometrics 121(1-2): 99-124.

Blundell, Richard, Lorraine Dearden, Alissa Goodman and Howard Reed. 2000. “The Returns to Higher Education in Britain: Evidence from a British Cohort.” Economic Journal 110: 82-99.

Boero Gianna, Abigail McKnight, Robin A. Naylor, and Jeremy Smith. 2004. “Graduates and the graduate labour market: evidence from the UK and Italy’” in Education, Training and Labour Market Outcomes in Europe, edited by Daniele Checchi and Claudio Lucifora. Palgrave Macmillan.

Bratti, Massimiliano, and Luca Mancini. 2003. “Differences in Early Occupational Earnings of UK Male Graduates by Degree Subject: Evidence from the 1980-1993 USR.” IZA Discussion Paper No. 890.

Brunello, Giorgio, and Lorenzo Cappellari. 2007. “The Labour Market Effects of Alma Mater. Evidence from Italy.” Economics of Education Review, forthcoming.

Bryson, Alex, Richard Dorsett and Susan Purdon. 2002. “The use of propensity score matching in the evaluation of active labour market policies.” Department for Work and Pensions Working Paper No. 4.

Bryson, Alex. 2002. “The Union Membership Wage Premium: An Analysis Using Propensity Score Matching.” Centre for Economic Performance Discussion Paper No. 530, London School of Economics.

Daymont, Thomas N., and Paul J. Andrisani. 1984. “Job Preferences, College Major and the Gender Gap in Earnings.” Journal of Human Resources 19(3): 408-428.

Dubin, Jeffrey A. and Daniel L. McFadden. 1984. “An Econometric Analysis of Residential Electric Appliance Holdings and Consumption.” Econometrica 52(2): 345-362.

Grogger, Jeff, and Eric Eide. 1995. “Changes in College Skills and the Rise in the College Wage Premium.” Journal of Human Resources 30(2): 280-310.

Haveman, Robert, and Barbara Wolfe. 1995. “The Determinants of Children's Attainments: A Review of Methods and Findings.” Journal of Economic Literature 33(4): 1829-1878.

Heckman, James J. 1979. “Sample Selection Bias as a Specification Error.” Econometrica 47(1): 153-161.

Imbens, Guido. 2000. “The Role of the Propensity Score in Estimating Dose-Response Functions.” Biometrika 87(3): 706-710.

James, Estelle, 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 79(2): 247-252.

Keane, Michael P. 1992. “A Note on Identification in the Multinomial Probit Model.” Journal of Business & Economic Statistics 10(2): 193-200.

Lechner, Michael. 2001. “Identification and estimation of causal effects of multiple treatments under the conditional independence assumption” in Econometric Evaluation of Labour Market Policies, edited by Michael Lechner, and Friedhelm Pfeiffer. Heidelberg: Physica-Verlag.

Lechner, Michael. 2002. “Some practical issues in the evaluation of heterogenous labour market programmes by matching methods.” Journal of the Royal Statistical Society A 127(1): 59-82.

Lee, Lung F. 1983. “Generalized Economic Methods with Selectivity Bias.” Econometrica 51(2): 507-512.

Loury, Linda D. 1997. “The gender gap among college-educated workers.” Industrial and Labor Relations Review 50(4): 580-593.

Loury, Linda D., and David Garman. 1995. “College Selectivity and Earnings.” Journal of Labor Economics 13(2): 289-308.

Makovec, Mattia. 2005. “Does it pay to study far from home? Explaining the returns to geographic mobility of Italian college graduates.” Manuscript, Bocconi University, Department of Economics.

McFadden, Daniel L. 1974. “Conditional Logit Analysis of Qualitative Choice Behavior” in Frontiers in Econometrics, edited by Paul Zarembka. New York: Academic Press.

Rochat, Denis, and Jean-Luc Demeulemeester. “Rational choice under unequal constraints: the example of Belgian higher education.” Economics of Education Review 20(1): 15-26.

Rosenbaum, Paul R., and Donald B. Rubin. 1983. “The Central Role of the Propensity Score in Observational Studies for Causal Effects.”' Biometrika 70(1): 41-55.

Sianesi, Barbara. 2004. “An Evaluation of the Swedish System of Active Labor Market Programs in the 1990s.” Review of Economics and Statistics 86(1): 133-155

Stewart, Mark. B. 1983. “On least squares estimation when the dependent variable is grouped.” Review of Economic Studies 50(4): 737-753.

Table 1: Summary Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Participation	34089	0.776	0.417	0	1
Full time employment	34089	0.604	0.489	0	1
Log(hourly net earnings)	22735	1.986	0.352	0.511	3.624
	College major				
Scientific	34089	0.231	0.421	0	1
Engineering	34089	0.215	0.411	0	1
Economics	34089	0.169	0.375	0	1
Social Sciences	34089	0.177	0.382	0	1
Humanities	34089	0.208	0.406	0	1
	High school type				
Vocational	34089	0.415	0.493	0	1
General	34089	0.172	0.377	0	1
Teaching/art	34089	0.306	0.461	0	1
Professional	34089	0.107	0.309	0	1
	Parents' education				
Both parents: elementary school	34089	0.119	0.324	0	1
At least one parent: junior high school	34089	0.098	0.297	0	1
Both parents:junior high school	34089	0.140	0.347	0	1
At least one parent: high school	34089	0.189	0.391	0	1
Both parents: high school	34089	0.197	0.398	0	1
At least one parent: university	34089	0.164	0.370	0	1
Both parents: university	34089	0.094	0.291	0	1
	Father's occupation				
Entrepreneur	34089	0.053	0.225	0	1
Professional	34089	0.073	0.261	0	1
Indipependent	34089	0.140	0.347	0	1
Other independent	34089	0.052	0.223	0	1
Manager	34089	0.102	0.302	0	1
White collar high level	34089	0.112	0.315	0	1
White collar low level	34089	0.194	0.395	0	1
Office worker	34089	0.104	0.306	0	1
Blue collar	34089	0.159	0.366	0	1
Other dependent	34089	0.021	0.143	0	1
	University Region				
Centre	34089	0.267	0.442	0	1
South	34089	0.264	0.441	0	1
North-west	34089	0.250	0.433	0	1
North-east	34089	0.219	0.413	0	1
	Age at the date of interview				
no more than 24	34089	0.053	0.224	0	1
25-26	34089	0.246	0.431	0	1
27-29	34089	0.459	0.498	0	1
more than 30	34089	0.242	0.428	0	1
Female	34089	0.537	0.499	0	1
High school score	34089	49.167	7.151	36	60
Military done before university	34089	0.032	0.175	0	1
Commuter within the same region	34089	0.590	0.491	0	1

Table 2: Evolution of graduates' composition by university groups

University Groups	2001	2004
Scientific	15.13	14.3
Engeneering	18.3	18.96
Economics	20.38	19.41
Social Sciences	23.82	25.11
Humanities	22.37	22.21
Number of obs	16,266	17,823

Table 3: Distribution of high school grades by university groups (%)

University Groups	High school final marks				
	36-40	41-45	46-50	51-55	56-60
Scientific	15.4	19.49	25.91	16.78	22.42
Engeneering	9.68	15.53	24.95	18.44	31.4
Economics	11.65	18.5	26.8	18.19	24.86
Social Sciences	16.32	20.94	26.76	15.56	20.42
Humanities	16.85	21.21	26.7	15.29	19.95
Total	14.14	19.29	26.29	16.74	24

Table 4: Distribution of high school types by university groups (%)

University Groups	High school types			
	scientific ghs	humanities ghs	vocational hs	other hs
Scientific	54.42	13.43	26.33	5.82
Engeneering	51.85	9.53	32.24	6.39
Economics	37.86	8.32	51	2.41
Social Sciences	29.89	32.91	27	10.08
Humanities	25.73	28.82	17	28.75
Total	38.24	19.89	30.46	11.4

Table 5: Labour market outcomes by university groups

University Groups	Log net hourly earnings (mean)	Participation rate (%)	Employment rate (%)
Scientific	2.002	74.65	79.07
Engeneering	2.049	89.26	88.17
Economics	2.005	81.36	87.32
Social Sciences	1.881	64.71	72.57
Humanities	1.938	74.48	59.37
Total	1.986	76.23	77.17

Table 5.1: Covariate balancing indicators before and after matching, outcome: participation

Treatment	N_1 Before	Comparison	N_0 Before	Median Bias Before (1)	Median Bias After (1)	N_1 off support
Scientific	4786	Engineering	5314	2.81	3.96	31
		Economics	4828	2.81	2.49	61
		Social sciences	5086	2.81	2.47	88
		Humanities	4840	2.81	1.93	39
Engineering	5314	Scientific	4786	2.26	1.68	134
		Economics	4828	2.26	2.6	197
		Social sciences	5086	2.26	2.02	194
		Humanities	4840	2.26	3.97	80
Economics	4828	Scientific	4786	5.26	2.84	74
		Engineering	5314	5.26	5.91	88
		Social sciences	5086	5.26	1.38	124
		Humanities	4840	5.26	3.07	99
Social sciences	5086	Scientific	4786	7	2.33	123
		Engineering	5314	7	2.52	188
		Economics	4828	7	2.84	197
		Humanities	4840	7	2.17	52
Humanities	4840	Scientific	4786	3.07	2.34	67
		Engineering	5314	3.07	5.11	98
		Economics	4828	3.07	2.08	64
		Social sciences	5086	3.07	2.89	199
with caliper=0.01						

Notes: N_1 indicates treated sample, while N_0 the non-treated one. N_1 off support indicates the number of observation not in the common support. Median absolute standardized bias before and after matching median taking all the regressors is calculated as follows: $B_{before}(X) = 100\frac{X_1 - X_0}{\sqrt{(V(X_1)+V(X_0)/2}}$, $B_{after}(X) = 100\frac{X_{1M} - X_{0M}}{\sqrt{(V(X_1)+V(X_0)/2}}$

Table 5.2: Covariate balancing indicators before and after matching, outcome: employment probability

Treatment	N_1 Before	Comparison	N_0 Before	Median Bias Before	Median Bias After	N_1 off support
				(1)	(1)	
Scientific	3611	Engineering	4770	3.27	3.31	51
		Economics	3923	3.27	2.52	85
		Social sciences	3379	3.27	2.3	140
		Humanities	3793	3.27	2.23	53
Engineering	4770	Scientific	3611	2.08	1.82	166
		Economics	3923	2.08	2.57	220
		Social sciences	3379	2.08	3.35	319
		Humanities	3793	2.08	3.36	101
Economics	3923	Scientific	3611	4.85	2.3	115
		Engineering	4770	4.85	5.08	72
		Social sciences	3379	4.85	2.92	186
		Humanities	3793	4.85	2.53	117
Social sciences	3379	Scientific	3611	5.55	3.19	216
		Engineering	4770	5.55	2.85	196
		Economics	3923	5.55	2.5	225
		Humanities	3793	5.55	1.86	53
Humanities	3793	Scientific	3611	2.66	4.04	143
		Engineering	4770	2.66	6.38	120
		Economics	3923	2.66	1.95	95
		Social sciences	3793	2.66	2.85	232
			with caliper=0.01			

Notes: see Note to Table 5.1

Table 5.3: Covariate balancing indicators before and after matching, outcome: hourly earnings

Treatment	N_1 Before	Comparison	N_0 Before	Median Bias Before (1)	Median Bias After (1)	N_1 off support
Scientific	2858	Engineering	4193	3.36	4.11	49
		Economics	3440	3.36	2.48	75
		Social sciences	2446	3.36	3.42	148
		Humanities	2267	3.36	3.98	42
Engineering	4193	Scientific	2858	2.66	2.74	165
		Economics	3440	2.66	1.93	154
		Social sciences	2446	2.66	2.72	327
		Humanities	2267	2.66	4.27	122
Economics	3440	Scientific	2858	5.32	2.2	152
		Engineering	4193	5.32	4.84	138
		Social sciences	2446	5.32	2.29	209
		Humanities	2267	5.32	2.57	139
Social sciences	2446	Scientific	2858	5.92	3.16	161
		Engineering	4193	5.92	2.96	163
		Economics	3440	5.92	3.05	186
		Humanities	2267	5.92	2.28	85
Humanities	2267	Scientific	2858	4.54	3.53	147
		Engineering	4193	4.54	6.17	174
		Economics	3440	4.54	2.91	103
		Social sciences	2446	4.54	2.54	100
with caliper=0.01						

Notes: see Note to Table 5.1

Table 6: Average treatment on the treated effects on the participation to the labour market

Treatment	Comparison	Participation			
		(1)		(2)	
Scientific	Engineering	-12.5	(-14.6; -10.2)	-12.4	(-15.2; -10.1)
	Economics	-5.4	(-7.9; -3.4)	-5.4	(-8;- 2.7)
	Social sciences	5.1	(2.5; 6.9)	5.6	(1.8; 7.5)
	Humanities	-3.6	(-6.6; -1.5)	-2.6	(-4.5; -0.4)
Engineering	Scientific	12.3	(10.4; 14.5)	12.0	(9.3; 14.8)
	Economics	8.8	(6.9; 11.3)	7.7	(5.7; 10.7)
	Social sciences	18.5	(15.8; 20.9)	16.6	(13.8; 19.5)
	Humanities	12.5	(9.3; 14.8)	11.2	(8.5; 14.2)
Economics	Scientific	4.6	(2.1; 6.8)	4.0	(11.9; 6.2)
	Engeneering	-6.4	(-8.7; -4.4)	-7.6	(-10.7; -5.1)
	Social sciences	10.2	(7.9; 12.6)	9.9	(7.2; 12.9)
	Humanities	4.1	(1.7; 7.4)	3.5	(0.4; 6.9)
Social Sciences	Scientific	-8.6	(-11.6; -6.9)	-6.6	(-9.8; -3.8)
	Engineering	-18.4	(-20.8; -15.4)	-17.7	(-20.7; -15.4)
	Economics	-13	(-15.4; -10.9)	-12.9	(-16.4; -10.7)
	Humanities	-10.7	(-13.4; -8.6)	-9.8	(-12.5; -7.3)
Humanities	Scientific	4.3	(24; 7.6)	3.6	(1.4; 7.8)
	Engeneering	-6.9	(-9.3; -4.1)	-7.3	(-9.9; -4.5)
	Economics	-4	(-7.7; -2.1)	-3.6	(-6.4; 0)
	Social sciences	10.2	(7.9; 12.7)	9.4	(6.8; 12.9)
	Caliper	*0.01*		*0.001*	

Note: Numbers in parentheses are the 95% confidence intervals based on White-corrected robust standard errors. 95% bias-corrected standard errors are bootstrapped using 500 replications. The average treatment effects are in relative terms expressed in percentage.

Table 7: Average treatment on the treated effects on employment rate

Treatment	Comparison	Employment			
		(1)		(2)	
Scientific	Engineering	-5.3	(-7.6; -3.2)	-4.8	(-7.7; -2.1)
	Economics	-7.8	(-9.8; -5.2)	-7.1	(-9.9; -4.1)
	Social sciences	4.8	(2; 7.4)	2.8	(-0.9; 6.8)
	Humanities	19.3	(16.4; 22.9)	20.2	(15.8; 24.2)
Engineering	Scientific	5.2	(2.6; 6.9)	4.2	(1.8; 7.6)
	Economics	-1.2	(-3.4; 0.4)	-0.7	(-3.1; 2.2)
	Social sciences	9.5	(6.8; 12.2)	9.7	(5.9; 12.9)
	Humanities	23.8	(17.3; 26.9)	24.9	(19.9; 29)
Economics	Scientific	6.5	(3.7; 8.5)	5.3	(2.5; 8.4)
	Engeneering	1.0	(-1.4; 3.3)	0.0	(-3.4; 2.4)
	Social sciences	10.0	(7.5; 13)	10.8	(5.5; 13.8)
	Humanities	25.9	(22.3; 29.2)	25.6	(21.3; 29.1)
Social Sciences	Scientific	-3.5	(-5.7; -0.1)	-2.4	(-6.2; 1.9)
	Engineering	-7.4	(-9.8;-4.1)	-8.1	(-11.1; -4.2)
	Economics	-13.9	(-16.3; -10.9)	-12.2	(-15.9; -8.6)
	Humanities	13.3	(9.2; 16.3)	13.8	(8.5; 16.7)
Humanities	Scientific	-16.4	(-18.9; -12.9)	-16.8	(-20.6; -12.9)
	Engeneering	-16.1	(-19.2; -11.5)	-18.7	(-21.7; -14)
	Economics	-26.4	(-30; -23;7)	-25.4	(-28.7; -21.7)
	Social sciences	-12.0	(-15.2; -9.4)	-13.5	(-17.3; -9.9)
	Caliper	*0.01*		*0.001*	

Note: see Note to Table 6

Table 8: Average treatment on the treated effects on hourly earnings

Treatment	Comparison	Hourly earnings			
		(1)		(2)	
Scientific	Engineering	1.0	(-1.1; 3.4)	0.6	(-2.8. 3.6)
	Economics	0.0	(-2.2; 2.2)	0.7	(-2.2; 3.4)
	Social sciences	11.0	(7.6; 13.3)	10.4	(6.7; 14.5)
	Humanities	5.6	(1.8; 8.2)	7.3	(3.3; 11.5)
Engineering	Scientific	1.1	(-0.9; 2.9)	0.3	(-2.6; 3.5)
	Economics	0.8	(-1.4; 2.9)	0.9	(-2.3; 3.3)
	Social sciences	15.2	(13.3; 19.1)	13.9	(10.5; 17.4)
	Humanities	8.3	(4; 12.5)	8.2	(4.3; 12.7)
Economics	Scientific	0.1	(-2.1; 2)	-1.1	(-3.7; 1.4)
	Engeneering	-3.4	(-5.8; -1.7)	-2.6	(-5.1; 0.4)
	Social sciences	12.5	(10; 15.5)	11.7	(7.8; 15.3)
	Humanities	6.4	(2.4; 9.4)	7.3	(3.2. 11.9)
Social Sciences	Scientific	-13.2	(-16.3; -10.8)	-11.9	(-15.6; -7.6)
	Engineering	-10.4	(-12.9; -7.2)	-13.0	(-16.8. -9.5)
	Economics	-12.4	(-14.9; -9.7)	-14.7	(-19.1; -11.1)
	Humanities	-6.5	(-9.2; -2.7)	-7.3	(-11.3;-3)
Humanities	Scientific	-5.5	(-8.2; -2.7)	-7.0	(-9.9; -1.8)
	Engeneering	-2.3	(-5.6; 1.7)	-1.8	(-5.1; 2.9)
	Economics	-5.1	(-8.9; -1.8)	-6.5	(-11.2. -3)
	Social sciences	6.3	(2.2; 8.2)	5.5	(0.8; 8.5)
	Caliper	*0.01*		*0.001*	

Note: see Note to Table 6

Table 9: Selection-corrected Log Hourly Earnings Equations Estimates (using the procedure suggested by Lee)

	Scientific		Engineering		Economics		Political Science		Humanities	
Variables	Coef	P-value	Coef	P-value	Coef	P-value	Coef	P-value	Coef	P-value
Parents' education:										
At least one parent: junior high school	0.020	0.399	0.010	0.642	0.016	0.484	0.017	0.645	-0.045	0.145
Both parents:junior high school	0.025	0.264	0.064	0.001	0.018	0.336	0.048	0.178	-0.002	0.949
At least one parent: high school	0.022	0.313	0.052	0.009	0.034	0.110	0.063	0.074	-0.013	0.625
Both parents: high school	0.054	0.021	0.068	0.001	0.017	0.443	0.086	0.021	-0.009	0.751
At least one parent: university	0.036	0.176	0.069	0.006	-0.008	0.798	0.071	0.101	0.004	0.895
Both parents: university	0.067	0.032	0.006	0.857	0.022	0.587	0.064	0.173	-0.002	0.959
Father's occupation:										
Entrepreneur	0.023	0.607	-0.006	0.848	0.076	0.032	0.030	0.522	0.060	0.228
Professional	0.042	0.257	-0.028	0.403	0.013	0.657	0.126	0.010	-0.037	0.408
Independent	0.012	0.666	-0.030	0.313	0.012	0.655	0.004	0.923	0.025	0.441
Manager	0.022	0.503	0.008	0.802	0.040	0.170	0.076	0.101	-0.019	0.714
Teacher	0.014	0.666	-0.033	0.288	0.027	0.342	0.026	0.549	0.020	0.594
White collar high level	0.032	0.258	-0.065	0.043	0.021	0.403	0.022	0.588	0.011	0.717
White collar low level	0.034	0.248	-0.013	0.658	0.025	0.356	-0.005	0.911	-0.018	0.590
Blue collar high level	0.012	0.664	-0.041	0.154	0.001	0.955	0.031	0.450	0.016	0.619
Female	-0.062	0.000	0.094	0.312	-0.101	0.000	-0.029	0.283	0.053	0.686
Region of residence:										
North-west	0.056	0.001	-0.079	0.074	0.132	0.000	0.076	0.155	0.055	0.025
North-east	0.008	0.644	-0.017	0.329	0.065	0.000	0.052	0.181	0.073	0.098
Centre	0.009	0.607	0.010	0.509	0.069	0.000	-0.009	0.705	-0.019	0.452
Age										
no more than 24	-0.087	0.012	0.368	0.018	0.037	0.506	0.099	0.001	-0.063	0.085
25-26	-0.057	0.012	0.215	0.015	0.012	0.732	0.038	0.310	-0.039	0.102
27-28-29	-0.033	0.042	0.097	0.012	0.002	0.951	0.080	0.000	-0.081	0.004
High school score	0.003	0.001	-0.007	0.153	0.004	0.000	-0.002	0.285	0.000	0.974
Survey year 2001	-0.133	0.000	0.048	0.325	-0.054	0.000	-0.102	0.000	-0.128	0.000
Correction term 1	0.049	0.622	-	-	-	-	-	-	-	-
Correction term 2	-	-	0.358	0.018	-	-	-	-	-	-
Correction term 3	-	-	-	-	0.007	0.943	-	-	-	-
Correction term 4	-	-	-	-	-	-	-0.365	0.320	-	-
Correction term 5	-	-	-	-	-	-	-	-	-0.212	0.308
Constant	2.028	0.000	2.567	0.000	1.866	0.000	1.403	0.000	1.669	0.000

Table 10: Results for the average treatment on the treated effects on hourly earnings: Lee model

Treatment	Comparison	Lee(ATT)	
Scientific	Engineering	-50.5	(-50.7; -50.2)
	Economics	4.6	(4.4; 4.8)
	Social sciences	49.7	(49.4; 50)
	Humanities	59.5	(59.3; 59.7)
Engineering	Scientific	42.9	(42.6; 43.4)
	Economics	46.3	(45.8; 46.8)
	Social sciences	82.1	(81.2; 83)
	Humanities	98.7	(98; 99.2)
Economics	Scientific	-3.3	(-3.5; -3.1)
	Engeneering	-52.6	(-52.9; -52.3)
	Social sciences	43.1	(42.7; 43.4)
	Humanities	54.7	(54.4; 54.9)
Social Sciences	Scientific	-48.9	(-49.1; -48.6)
	Engineering	-98.9	(-99.4; -98.5)
	Economics	-45.7	(-45.9; -45.5)
	Humanities	5.6	(5.3; 5.8)
Humanities	Scientific	-47.3	(-47.8; -46;7)
	Engeneering	-92.5	(-93.3; -91.6)
	Economics	-44.2	(-44.8; -43.7)
	Social sciences	-4.3	(-4.5; -4.1)

Note: : Numbers in parentheses are the 95% confidence intervals based on White-corrected robust standard errors. Standard errors take into account for the estimated coefficients used to construct the conditional means and are therefore precise. Average treatment effects are in relative terms expressed in percentage.

Table 11: Results for the average treatment on the treated effects on hourly earnings: Binomial logit estimates

Treatment	Comparison	PMS(ATT)		PMS(ATT)	
Scientific	Engineering	-0.4	(3.4; 1.7)	0	(-2.03; 2.37)
	Economics	0.8	(-1.0; 2.6)	0.8	(-1.53; 2.43)
	Social sciences	12.3	(9.1; 15)	12.2	(7.78; 14.57)
	Humanities	4.6	(0.2; 7.5)	4.6	(1.48; 7.09)
Engineering	Scientific	1.3	(-0.2; 4.2)	1.4	(-0.32; 3.67)
	Economics	-0.7	(-3; 0.7)	-0.7	(-3.75; 0.64)
	Social sciences	11.2	(7.8; 13.7)	12.3	(8.67; 14.17)
	Humanities	4	(0.7; 8.8)	4.1	(-0.59; 7.61)
Economics	Scientific	-0.8	(-2.9; 1.2)	-0.5	(-2.06; 1.68)
	Engineering	-1.5	(-4; 0.3)	-1.2	(-3.77; 0.59)
	Social sciences	13.5	(10.5; 16.2)	13	(10.78; 16.51)
	Humanities	5.5	(2.2; 8.7)	5.2	(2.42; 8.57)
Social Sciences	Scientific	-14.7	(-17.3; -12.3)	-14.6	(-17.12; -12.31)
	Engineering	-12.1	(-15.1; -9.5)	-12.7	(-15.57; -10.13)
	Economics	-11.6	(-13.9; -9)	-11.7	(14.50; -9.26)
	Humanities	-8.3	(-12.5; -5.2)	8.6	(-11.65; -6.26)
Humanities	Scientific	-5.5	(-7.5; -3.1)	-5.5	(-7.88. -2.39)
	Engineering	-4.5	(-7.8; -1.4)	-4.2	(-6.88; -0.37)
	Economics	-2.7	(-4.7; 0)	-2.6	(-4.50; 0.62)
	Social sciences	7.2	(4.2; 10.4)	7	(3.93; 10.17)
	Caliper	*0.01*		*0.001*	

Note: Bold type indicates statistical significance at 5% level. Results are in relative terms expressed in percentage.

Figure 1: Log net hourly earnings.

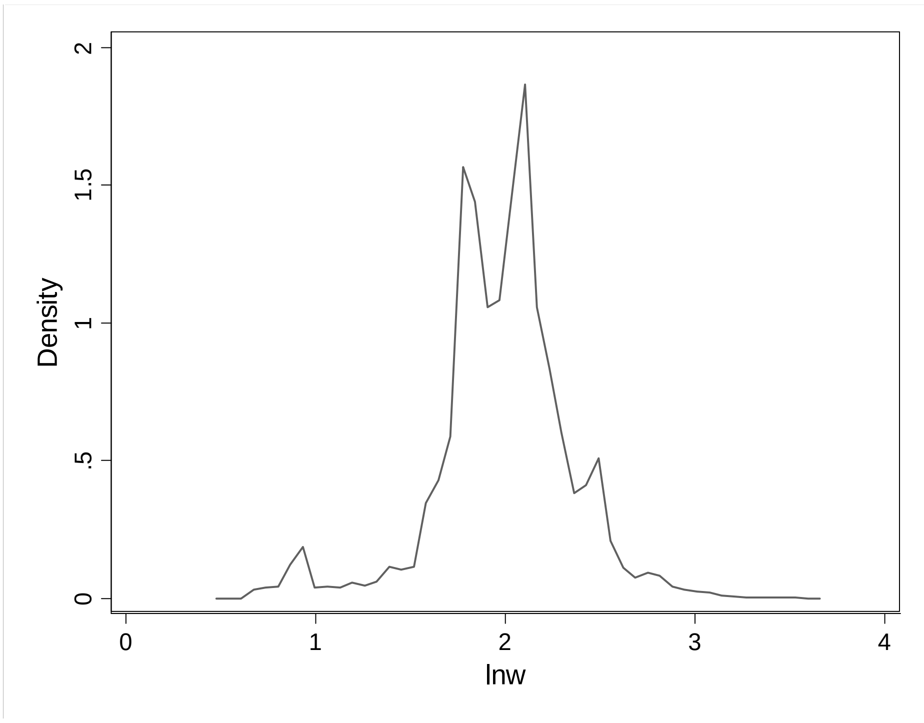
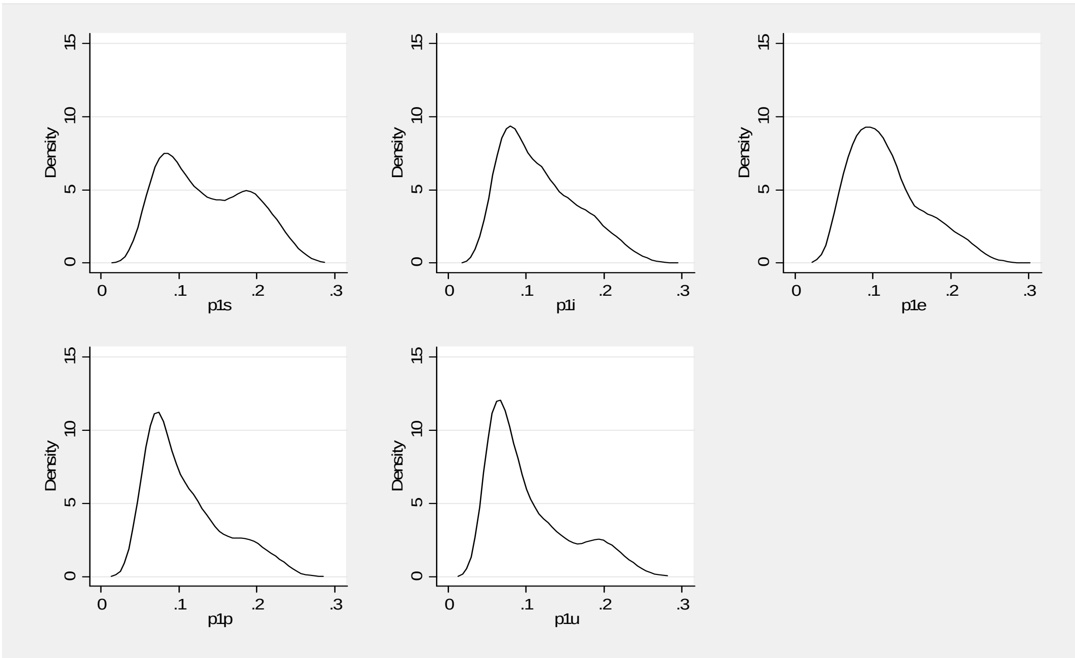
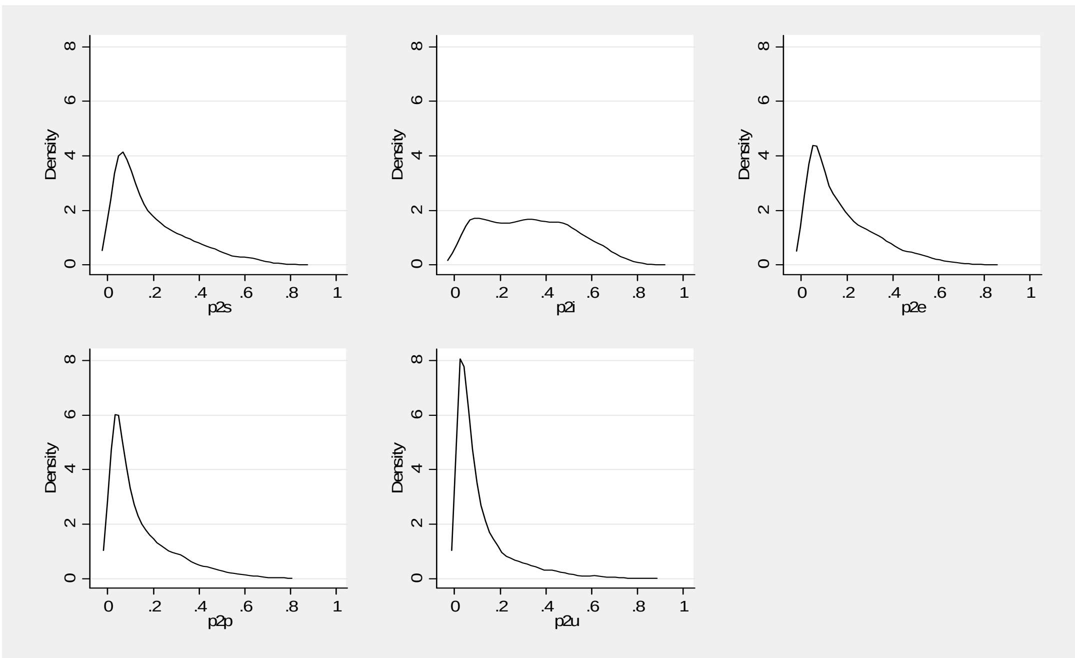


Figure 2: Distributions of the estimated propensities to be assigned into the fields of study. Sample of all university graduates.

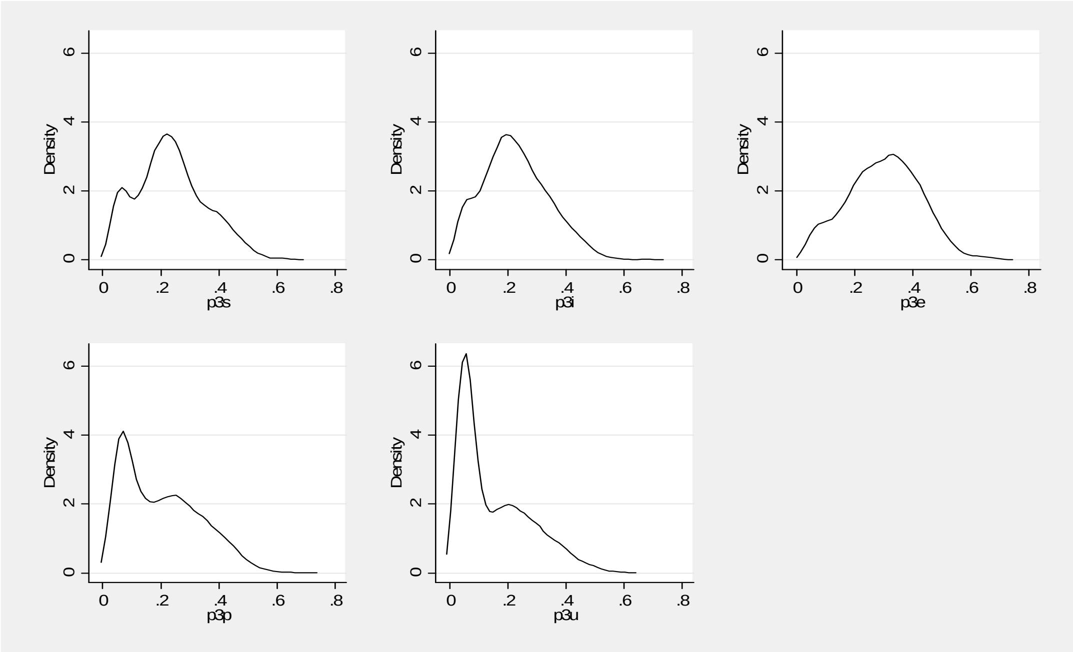
Estimates for propensity to enrol at the Scientific Field.



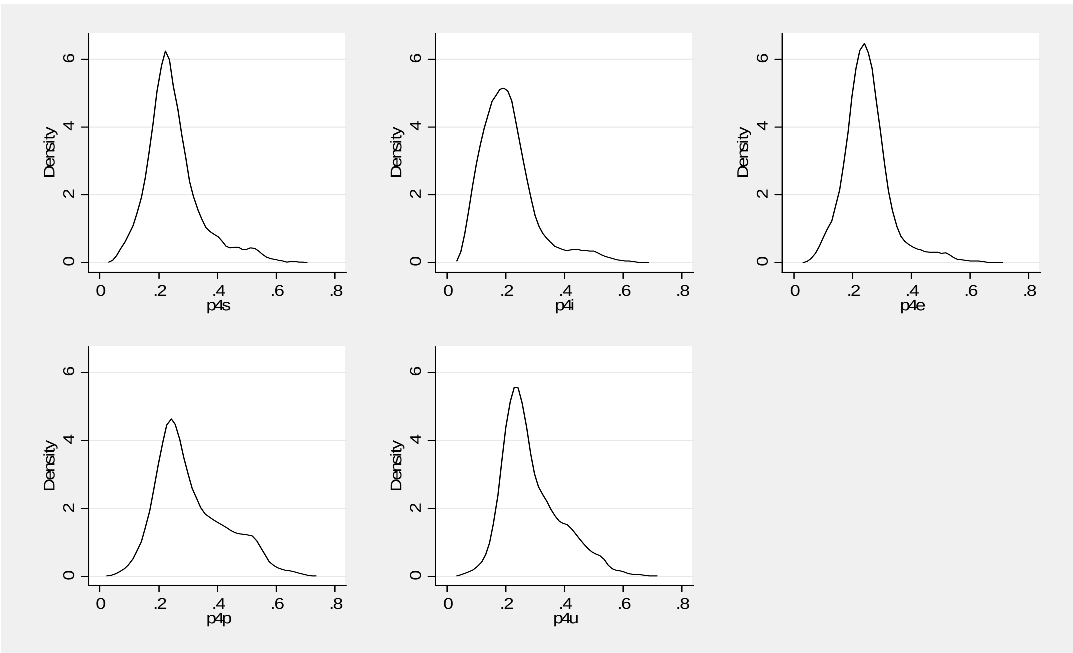
Estimates for propensity to enrol at Engineering.



Estimates for propensity to enrol at Economics.



Estimates for propensity to enrol at Social Sciences.



Estimates for propensity to enrol at Humanities.