Returns to education on earnings in Italy: an evaluation using propensity score techniques

Elena Cottini

Università Cattolica del Sacro Cuore, Milan

PRELIMINARY DRAFT

(June 2005)

ABSTRACT:

This paper investigates the returns to different levels of education using a matching estimator and compares these results to those obtained with more classical estimation methods like OLS. Empirical evidence is based on microdata on the 1998 cohort of Italian high school graduates observed three years after graduation. In the empirical analysis we assess the earning differential of university versus high school education taking into account drop out students. Probit estimates show that parental level of education, academic performance and general high school influence positively the probability of obtaining a university degree. Then a matching estimator is used to grange the effect of school achievement on earnings. Results suggest that controlling for ability and eliminating selection bias high education has a positive impact on earnings. This conclusion leads to important policy implications both for the financing of the different levels of education but also on individual decisions on types of degree.

Key words: returns to education, treatment effects, propensity score matching

JEL CODES:J38, J65, J68

1. Introduction

University education is generally seen as a prerequisite for labour market success. In a period where the financial costs to students of acquiring higher education are rising it is important to understand how the benefits from such an investment may differ. In this paper we focus on the Italian situation: the Italian school system is based on primary, secondary and higher education. The primary level starts at the age of six until the age of 11, then follows secondary school untill 14, that is the age when compulsory education is completed. At the end of compulsory education children can decide to enter higher education or work, in the first case the diploma obtained at the end of high school gives access to professions or to University. The Italian high school system consists of three types of high schools: liceo (general education), istituto tecnico or professionale (that gives a technical or professional education) and finally istituto magistrale (that consists in teaching schools). The first two types of institute last five year and afterwards students can decide to enrol at University or go to work, while the third type -typically for those students who desire to become teachers- lasts four years plus an additional year for those students who wish to go to University. Literature on education asserts that completing a general secondary school (liceo) without going on with enrolment at the University yields a very limited return. For example Checchi (2000) finds that attending a professional (istituto tecnico) or a vocational (istituto professionale) school yields a yearly rate of return of about 6%, whereas attending a general school (liceo classico or liceo scientifico) generates a lower return of 5% for every additional years of education achieved. However, if a student decides to continue with the university, the rate of return raises significantly, from the lowest 7.3% for Literature and Philosophy, passing through 11.7% for Economics and Political Sciences up to 13.9% for Law. This situation suggests that the educational career is significantly predetermined by the choice of the secondary school undertaken at the age of 14. This decision is generally influenced by families, in fact, those parents able to finance a university education for their children and/or expecting them to be above a minimum level of performance tend to push them to register in general schools (licei) even if these schools offer a lower rate of return per se. The return, once finished university, will be much higher. Ideally the way we would wish to measure the return to schooling would be to compare the earnings of an individual with two different levels of schooling, but in practice only one level of education is observed for a particular individual. The literature has recently attempted to deal with this problem by finding 'experiments' in the economy that randomly assign groups of individuals to different levels of schooling. In this paper matching methods are used to represent, depending on the particular method employed, either a semi-parametric or a non-parametric alternative to linear regression (Black and Smith, 2003)[1]. In fact linear regression approach has some limitations: the first consists in the fact that the linearity assumption can hide the failure of the "common support" condition, while the second is the selection bias that is not taken into account using linearity and conditioning only on observables. To explain the situation, consider the case in which only high grade students attend university and only low grade students go to work. The counterfactual outcome -what high grade students experience when going to work- is not non-parametrically identified. Instead, the linear functional form assumption does not consider this problem and always identifies the counterfactual outcome. While matching does not solve the common support problem, it reduces selection bias, something that linear regression does not. In order to reduce the "dimensionality" of our matching problem, we employ propensity score matching methods, thanks to which we match individuals on the predicted probability of attending university, which is a function of observed X, rather than matching directly on X. Once we have the distributions of estimated propensity scores for sample individuals in universities and working activity, we can compare the two densities to get a clear sense of the extent of the common support problem. By constructing an observation-specific counterfactual for each treated observation, matching methods avoid bias due to misspecification of the functional form in a linear model. Untreated observations similar to each treated observation in

[1] See the discussions in Heckman, Ichimura, and Todd (1997,1998), Heckman, Ichimura, Smith, and Todd (1998), Heckman, LaLonde, and Smith (1999), Dehejia and Wahba (1999,2002), and Smith and Todd (2003). Matching methods require binary treatments.

terms of the probability of participation, P(X), serve as counterfactuals.

Then we compare the estimates from propensity score matching to estimates of the same parameter based on the standard linear regression specification in the literature. Finally we consider a multiple treatment model, which distinguishes the separate impact of different levels of education on earnings. The data set used by most studies on Italian returns on education is the *Bank of Italy Survey on Household Income and Wealth* which is cross section and contains data only on the highest degree attained at school. In this paper is used an alternative and much more informative dataset derived from *A survey on the transition from high school to work or university for high school graduates of 1998* collected by the Italian National Statistical Office (ISTAT). More detailed information on data are contained in section 7. Our aim is to examine how students of different abilities and background sort into university. Therefore the experience of a sample of high school students who graduated in 1998 in terms of their choice of enrolling at university or going to work could help to inform next generations and policy makers. Moreover, given the high rate of students who abandon university in the first years, the role of drop out is analysed [2]. The paper is organised as follows. Section 2 presents an essential review on the empirical literature on returns to education. In section 3 the causality problem between education and labour market outcomes is described. The evaluation problem and theoretical set up of matching estimator is presented in section 4.The identification problem and propensity score technique is exposed in section 5, and in section 6 data and some descriptive statistics are given. Section 7 shows the results from propensity score matching approach and finally section 8 concludes. Figures and Tables are presented in Appendix C. Appendix A presents a list of all variables used in estimation while Appendix B the estimation results of dependent variable[3]. Figures and Tables are presented in Appendix C.

[2] In our context we want to compare three different groups of high school leavers: drop out , workers , students three years after graduating from high school.

[3] Our dependent variable is log (earning) that is banded then we rebuilt a continue variable.

2. Review of the empirical evidence (to complete)

The estimation of the returns to schooling, both for the individual and society more generally, has been the focus of considerable debate in the economics literature. Different methods have been used to quantify the returns to schooling like OLS, instrumental variables and experimental methods. Angrist and Krueger (1991) compare Ols to IV results using the presence of compulsory schooling law variation across US states and the quarter of the year in which a person was born as the basis of their instruments. The underlying idea here is that a person who has been born early in the first quarter of the year reaches the minimum school leaving age after a smaller amount of schooling than persons born later in the year. The actual amount of schooling attained is directly related to the quarter in which they were born while at the same time there seems no reason to believe that quarter of birth has an own independent effect on earnings. Direct estimation by OLS gives an estimate of the return to schooling of 0.063 whereas the IV method gives an estimate of 0.08119. Card (1995) uses an indicator for the distance to college as an instrument for schooling based on the observed higher education levels of men who were raised near a four-year college and finds returns of 13.2% compared to OLS estimates of closer to 7%. A somewhat different approach is used in the paper by Duflo (1999) where estimation is based on the exposure of individuals to a massive investment program in education in Indonesia in the early 1970's. Individuals were assigned to the treatment on the basis of their date of birth (pre and post reform) and the district they lived in (as investment was a function of local level needs assessment). A sort of natural experiment was performed. The paper by Card (1999) is the most recent comprehensive study, which compares OLS, matching and IV estimators in models with heterogeneous treatment effects. Harmon and Walker (1995) exploit the natural experiment of a change in the minimum school-leaving age to circumvent the need to observe ability and family background variables. Dearden (1999a and 21999b) and Blundell, Dearden, Goodman and Reed (2000) both use the British NCD cohort data. Overall, for the UK, most authors choose to adopt qualification-based measures of educational attainment rather than years of education. Finally Blundell, Dearden and Sianesi (2004) use different approaches for recovering the impact of education on individual earnings both for single treatment and sequential multiple treatments with and without heterogeneous returns. They show the importance of correcting for detailed test score and family background differences and of allowing for (observable) heterogeneity in returns. And finally they find a return of 27% of higher education versus anything.

3. Modelling Returns to education: homogeneous versus heterogeneous returns

In early literature on returns on education the homogeneous approach has been widely used while in recent works authors focused on models that allow for heterogeneous returns (Heckman, Smith and Clemens (1997), Blundell, Dearden and Sianesi (2002)). Let us consider a general model of returns to education:

$$Y_i = \alpha + \beta_{ij} \sum_j S_i + \delta X_i + \varepsilon_i$$

where α represents the relative level of earnings across individuals for any given level of schooling and β represents the impact of schooling level *j* relative to the base level. We have homogeneous returns if $\beta_{ij} = \beta_j$ for all individuals *i*, while returns are heterogeneous when this condition does not hold. Homogeneous returns have been estimated using Ordinary least squares methodology while for heterogeneous returns we use matching and interacted Ols estimations.

3.1. Econometric Methodology

This section describes three different methodologies to estimate high school's graduates earning differentials: i) ordinary least square estimation (OLS); ii) Interacted Ordinary least squares estimation (IOLS) and iii) propensity score matching-average treatment on the treated method (ATT).

3.1.1 Least Squares

A widely used method to quantify returns to education is to estimate a model as:

$$Y_i = \alpha + \beta S_i + \delta X_i + \varepsilon_i \qquad (2)$$

where Y_i is the natural logarithm of earnings for individual i, S_i is a dummy vector that assumes value 1 if the individual continues studying after graduating from high school and value 0 if he/she starts working, X_i is a vector of individual characteristics that might affect both choice of continue studying and occupational earnings, β is the average effect of higher education on earnings compared to the state of low education.The error term is assumed to be independently and identically distributed across individuals with $E(\varepsilon_I)=0$. In this case we assume <u>homogeneous returns</u> to higher education this implies that the Average treatment Effect (ATE), the Average Treatment on the Treated (ATT) and the Average treatment on the Untreated (ATNT) all coincide with estimated β's. We can assume that since the level of education obtained and educational choices seem to depend on expectations on future earnings, the education dummy S_i is correlated with b's. When we allow for <u>heterogeneous returns</u> to education we need to use fully Interacted Ols or matching estimation methods[4]. Let's start with fully interacted Ols:

$$Y_i = \alpha + \beta S_i + \gamma_1 (X_{1i} S_i) + \gamma_2 (X_{2i} S_i) + \gamma_3 (X_{3i} S_i) + \delta_1 X_{1i} + \delta_2 X_{2i} + \delta_3 X_{3i} + .. + \varepsilon_i$$

where ATT≠ATE≠ATNT :

$$ATT = \delta + \delta_1 E(X_1 \mid S = 1) + \delta_2 E(X_2 \mid S = 1)$$
$$ATE = \delta + \delta_1 E(X_1) + \delta_2 E(X_2)$$
$$ATNT = \delta + \delta_1 E(X_1 \mid S = 0) + \delta_2 E(X_2 \mid S = 0)$$

The statistical significance of the interaction terms provides evidence of the presence of heterogeneous returns. Finally we should take into account that performing OLS estimation allows for different sources of bias due to unobservables (Blundell, Dearden and Sianesi (2003)):

a) Ability bias: this derives from correlation between intercept term and the scholing dummy:we might expect that high ability individuals acquire both higher education and higher earnings. The observer can't observe the ability of individual. This generates an unpward estimate of ATT and ATE.

b) Return Bias: this happens when the individual's returns component β is correlated with the decision of which level of education to achieve. In this case returns to higher education might be not homogeneous across individuals[5] and the estimates will be biased with respect to ATE.

c) Measurement error bias: educational variables might be measured with errors, since the level of education is reported by a dummy variable the measurement error will vary with the level of education achieved and it is not classical.

Using treatment matching estimator and interacted Ols method ability bias and return bias are reduced (Blundell et all (2003)), while the third source of bias should diminish considering educational qualification rather than, like in great part of education literature, years of schooling (Bratti, Naylor and Smith (2005)).

3.1.2. Matching

Matching method is a non parametric procedure to identify the impact of a treatment on an outcome. The framework that guides this approach is the potential outcome approach to causality suggested, in its first version, by Roy (1951) and Rubun (1974). Let us conceive of " achieving higher education" (HE) as if it was a "treatment" that the individual receive; and we would like to evaluate the causal effect of this treatment relative to the other possible situations (NHE). Then, we consider two population of individuals and for each population measure two variables D and Y. We want to evaluate the effect of a treatment D[6] on Y, called outcome of the ith. Moreover denote variables unaffected by the treatment, called covariates or pre-treatment characteristics by X. Suppose that the variable D could assumes two values:

[5] In the homogeneous returns model the second source of bias is obviously absent

[6] In our case is attending university but could be a lot of different tings such as an additional year of education, a course of specialization, a medical treatment or whatever that implies the observation of a pre-treatment population and a post-treatment population.

D=1 if the individual i was exposed to the treatment

D=0 if the individual i was not exposed to the treatment

both situations are possible before the treatment, while if the treatment takes place only one of the two occurs because the ith individual cannot be contemporaneously treated and not treated. Then if the individual is treated we define counterfactual the situation in absence of the treatment. To indicate the causal relation between the outcome Y and the treatment D we use the notation $Y_i(D_i)$, finally we can describe the two events that are related to the ith individual as:

Y =(1) is the outcome if treated

Y =(0) is the outcome if not treated

We can at this point give a definition of A CAUSAL EFFECT:

The treatment D has a causal effect on the outcome Y for the ith individual if the result in case of treatment is different from the result in absence of the treatment, that means

$$\Delta_i \equiv Y_1 - Y_0 \neq 0$$

in this case Δ_i represents the causal effect of D on Y for the ith individual. This relation sheds light on how problematic can a causal relation be, for example if our aim is to analyse the effect of a higher level of education on one individual's earning, it is not sufficient to study the direct effect on the individual's earning but also what would have happened if the individual had not achieved a higher degree (Ichino,1999).

The evaluation problem is a problem of missing data (Larsson, 2000) since we cannot observe the counterfactual, referred as the outcome which would have resulted if an individual had made an alternative choice (that is to say if individuals had chosen to go to work and *vice versa*). Thus the true causal effect of a treatment T on individuals not subjected to the treatment can never be identified, because the counterfactual cannot be inferred directly from the outcomes of working individuals since they are likely to differ substantially in their characteristics from those who attend university. In order to identify individual treatment effects it is important to make some assumption on the distribution of Y_1 and Y_0. The commonly estimated parameter of interest is the mean impact of the treatment on the treated:

$$\begin{aligned} ATT &= E(Y_1 - Y_0 \mid D = 1, X) = \\ &= E(Y_1 \mid D = 1, X) - E(Y_0 \mid D = 1, X) \end{aligned}$$ [7]

where D=1 denotes treatment and D=0 denotes non-treatment and X is a set of conditioning pre-Treatment characteristics. The parameters of interest in such analysis are three: Average Treatmentr Effect (ATE), Average Treatment on the treated (ATT) and Average Treatment on the Unttreated (ATNT). In calculating ATT we are taking into consideration how are individuals earnings compared with what they would have been if they had not been starting university, on average. For university students we observe Y_1[8] so that the average observed outcome for university enrolled students is an unbiased estimate of the first component of the second equation. Problems come out with the second term of the equation, which represents the mean of the counterfactual that is unobserved and consequently must be estimated thanks to the untestable identifying assumption that forces us to use the observable pairs (Y_1,D=1) and (Y_0,D=0). Thus, E(Y_1 |D=1)-E(Y_0 | D=1) would be in general biased because of the effect of the treatment on the treated. This situation is overcame when Y_0 is independent of D, which occurs when the random assignment ensures that potential outcomes are independent of treatment status. In this situation occurs that:

$$E(Y_0 \mid D = 1) = E(Y_0 \mid D = 0) = E(Y \mid D = 0)$$

7 The average treatment effect on the non-treated $ATNT = E(Y_1 - Y_0 \mid D = 0) = E(Y_1 \mid D = 0, X) - E(Y_0 \mid D = 0, X)$.
While finally the average treatment effect is: $ATE = E(Y_1 - Y_0)$ is a weighted average of the treatment effect for the treated and the non treated.

8 We actually do not have earnings of those individuals who attend university and we are forced to estimate these starting from the β's of university graduates regressed on characteristics of the former. Further details are presented in Appendix D.

then the treatment effect can be consistently estimated by the difference between the observed mean of the outcome variable for the treatment group and the observed mean for the non treated group.

However the average causal effects described by the equation below can be identified if a particular condition is satisfied[9]. The application of these methods is quite difficult to perform in economic analysis, consequently- for example- to ethic problems. These problems rise when, for example, individuals choose whether to take part or not into a programme (where with programme we refer to achieving a higher degree, taking part into a training programme, assuming a new treatment or taking part into a professional courses, etc..). If random assignment to the treatment is not possible a possible solution is to construct a comparison group, that apart from not being in the treatment, is much similar as possible to the treated group.

To overcome the selection problem described above researchers could choose from a range of evaluation methods, and this choice is determined by a number of factors including the richness of the data and the nature of the treatment . In experiments the participants are randomly assigned to the treatment from a large group of eligible possible participants. Then, in a binary case, we obtain an unbiased estimate of the average treatment effect comparing the treated and the control group. Hence the aim of a non-experimental evaluation is to construct a comparison group that is as much similar as possible to group exposed to the treatment . One method proposed to solve this problem is matching.

Matching methods have been widely used in statistics and medical literature both in statistics and medical literature, and in empirical and theorical works (Rosenbaum and Rubin (1989), Heckman, Ichimura and Todd (1998), Dehejia and Wahba (1999), Abadie and Imbens (2002), Hirano and Imbens (2002)). But only recently they have been applied to labour economics (Larsson (1999), Lechner (1999), Martin (1998), Hardoy (2000), Ichino and Nannicini (2004) and Sianesi (2001 and 2003)) to evaluate active labour market policies and returns to education in UK (Sianesi, (2005)).

9 This condition is the Conditional Independency Assumption (CIA): all differences affecting the selection between the group of participants in the treatment and the group of non-participants are captured by observable characteristics. The CIA will be presented in the following section.

This method has been efficiently applied in presence of two situations: (a) the presence of a large number of controls and (b) the interest in analysing average treatment effect on the treated group of individuals (Imbens 2003). In short this method involves in pairing together individuals from the treated group to individuals from the control group who are similar in terms of their observable characteristics and selection into the treatment is exclusively carried out taking into account observable pre treatment characteristics, then matching on them yields unbiased estimates of the average treatment effect. Finally it is based on the intuition of contrasting the outcomes of programmes participants (indicated with Y_1) with the outcomes of comparable non-participants (indicated with Y_0) and the differences between the two groups are solely attributed to the treatment. The method of matching is intuitively appealing and is difficult to implement, consequently it is often used by applied statisticians rather than economists (Heckman, Hichimura and Todd, 1997). First of all it is difficult to determine if a particular comparison group is comparable to the treatment group because we need them to having obtained the same outcomes as the treatment group if they had participated into the programme, but matching on the available characteristics in a typical non experimental study does not guarantees to produce such a control group. That is to say that this method estimates the effect of the treatment within a group of individuals, having similar characteristics, as a difference between the outcomes referred to the two groups of individuals, with an estimator for the Average Treatment on the Treated (ATT) that is obtained by averaging these within pair-differences. Matching estimators try to resemble an experiment by choosing a comparison group from all non-participants such that the selected group is as similar as possible to the treatment group in observable characteristics. Matching can yield unbiased estimates of the treatment impact where differences between individuals affecting the outcome of interest are captured in their observed attributes. This assumption, which is often referred to as Conditional Independence Assumption (CIA) is the key identifying assumption underlying matching methodology.Naturally the precise form of the CIA depends on the parameters being estimated. Formally it can be expressed as:

$$E(Y_0 \mid X, D=1) = E(Y_0 \mid X, D=0)$$

Thus, CIA requires that the chosen group of matched controls does not differ from the group of treated by any variable that is systematically linked to the non participation outcome Y_0 other than on those variables that are used to match them. For the treatment on the treated parameter, the CIA requires that, conditional on observable characteristics, potential non treatment outcomes are independent on treatment participation. But this assumption is often not consistent with great part of economic models where agents select into the programme on the basis of unmeasured components of outcomes unobserved by the econometrician. Even if CIA is achieved for one set of control variables, it is not guaranteed to be achieved for other sets of variables including those that include the original variables as subsets.

Secondly, if a valid comparison group is found the distribution theory for the matching estimator remains to be established for continuously distributed matched control variables. Third, most of the current econometric literature is based on separability between observables and unobservables and on exclusion restrictions that isolate different variables that determine outcomes and programme participation. Separability permits the definition of parameters that do not depend on unobservables. Exclusion restrictions arise naturally in economic models, especially in dynamic models where the date of enrolment into the programme differes from the date when consequences of the programme are measured. Under CIA we have that the mean of the potential not treated outcome is the same for those receiving treatment as for those not receiving treatment. This permits to use non participants outcomes to infer participants counterfactual outcomes. But this is valid only if there are non participants for all participant's values of X. This is the common support condition, that can be exposed as:

$$0<\Pr(D=1|X)<1$$

If there are regions where this condition is not satisfied, the support of X does not overlap for the treated and non-treated groups, matching can be performed and the treatment calculated in the common support,

while if treated individuals have no support in the non treated population they are dropped from analysis.
Finally matching needs a huge dataset, in fact with a large number of conditioning variables it is easy to have many cells without matches, this makes the method depending on arbitrary sorting schemes to select hierarchies of matching variables (Westat ,1982). But we must underline that Rosembaum and Rubin (1983) partially solve this problem establishing that if matching on control variables is valid it would be valid also on the probability of selection into the programme $Pr(D=1|X)=P(X)$.
They propose a method that adjusts for pre treatment characteristics based on the conditional probability of receiving the treatment given pre treatment variables, dimostrating that solely adjusting for the propensity score removes all the bias associated with differences in pre-treatment variables in treatment and control groups.
They conclude that if the CIA is valid for X, it is also valid for a function of X called the balancing score b(X).The advantage of the balancing score property is the decrease in dimensionality (Larsson,1999), in fact, instead of conditioning on all covariates it is sufficient to condition on some function of the covariates, rather than matching on a vector of characteristics, it is possible to match on just the propensity score. This is because, Rosembaum and Rubin show, that treatment and non treatment observations with the same value of propensity score have the same distribution of the full vector of regressors.
A generalized version of the CIA assumption is in literature known as the conditional Bias Stability Assumption (BSA). This identifying assumption has the big advantage of containing the CIA and it allows to test its validity (Eichler and Lechner, 2001). Furthermore, assuming the BSA requires less information for ATT identification than under the CIA. If the CIA is satisfied matching is an attractive method to evaluated the impact of university education on earnings for two main reasons: first it is non parametric avoiding the need to define a specific form for the outcome equation, second it avoids extrapolation beyond the common support which occurs with simple linear regression.

The most common matching estimator is the propensity score matching estimator proposed by Rosenbaum and Rubin (1983), who showed that, conditional on the propensity score P(X), the treatment and the observables are independent (balancing property):

$$D \perp X \mid P(X)$$

Identification relies on the CIA or Unconfoundness: conditional on the set of observable X (or P(X)), the non-treatment outcome is unrelated to the treatment:

$$Y_1, Y_0 \perp D \mid P(X)$$

The treated and the untreated can then be matched on the basis of their propensity score P(X). The identification and estimation of the average treatment effect for the whole population can be developed in several ways, Lechner (1999), for example, suggests that the average treatment effect on the population is identified by a weighted sum of the treatment effect on the two sub-samples. For more detailed information on the identification issue see Lechner (1999) and Imbens (1999).
Since the estimated P(X) is a continuous variable and matching on exactly the same value of P(X) is practically unfeasible, several matching procedures have been developed in literature and they differ in terms of the system of weights adopted for choosing the potential controls (such as assigning a unity weight to the nearest untreated observations and zero to all the others; equal weight to all the controls within a certain radius from the propensity score of the treated; kernel weights; etc.).The proper use of Propensity Score Matching estimators allows then to reduce (not to eliminate) the evaluation bias, furthermore availability of a rich data set should make this type of bias less relevant. More in general, the reliability of the results obtained with any matching estimators heavily depends on the quality of data.
Detailed information on the pre-treatment characteristics of both the treated and the untreated is actually crucial to make the CIA assumption convincing. Despite of the quality of the information, the CIA is a quite strong assumption if the individuals decide also in the basis of their forecast outcome (Blundell and Costa Dias, 2002).

Rosembaum and Rubin's approach deals only with a treatment that takes two values, but a treatment might take more than two values. Single treatment's approch has been recently extended to the case of multi-valued treatments by Lechner (2001) and Imbens (2000) mantaining the avantages of propensity score approach. Like in the binary case this methodology requires the estimation of a conditional expectation of the outcome of interest as a function of one variable for each level of the treatment and it relies on a weaker version of the unconfoudedness assumption[10]. Finally, similarly to binary treatment, multiple treatment techniques need to estimate the propensity score. With a whole range of treatments available two cases of interest can be distinguished: first, the case in which the treatments are not logically ordered; second, the treatments are logically ordered. This second case is our case, and the propensity score is estimated using ordered probit model.

4. Description of the dataset:summary statistics

The survey used in this paper was carried out by ISTAT, the Italian National Statistical Office, on a regular basis and it is on the transition from high school to work or university of a representative sample of Italian high school leavers. The wave used covers individuals who graduated in 1998 and were interviewed in 2001, three years after completion of the degree. The sample who answers the questionnaire corresponds approximatly to the 5% (23262) of the population of high school leavers of 1998 (478904). The questionnaire is composed of three main sections: the first contains questions on student curricula and training, the second on job searching and job characteristics and the last on family background and individual characteristics[11].

[10] It relaxes two aspects of strong unconfoundedness: first it requires only pairwise independence of the treatment with each of the potential outcomes; second, it requires the independence of the potential outcome and the tteatment to be local at the treatment level of interest.

[11] Then more precisely there are information on school curricula, university studies, post-high school training labour market experience in the three years after graduation, interruption of University studies, job search activities, parental education and occupation, composition of the household and individual information (age, zone of residence, sex, nationality).

Sample means are presented in Table 4[12] and a more detailed description of the variables used in the estimations is reported in Appendix A. The Italian high school system at the time of the interview consists of three types of high schools: liceo (general education), istituto tecnico or professionale (that gives a technical or professional education) and finally istituto magistrale (that consists in teaching schools). The first two types of institute last five year and afterwards students can decide to enrol at University or to work, while the third type --typically for those students who desire to become teachers- lasts four years plus an additional year for those students who wish to go to University. From Fig.1 we derive that graduates from general high school represent nearly 19% of the whole sample, while graduate from technical/professional institutes constitute the 73% and finally those who exit from teaching schools consist in only the 7%.

The relation between parental education and high school types is described in Table 1, where we see that the more educated are the parents the more likely are children to get general education. There can be three possible explanations to this situation according to Cappellari (2003). One is referred to preferences, as long as more educated parents give higher value to education and prefer to enrol their children in general institutes which would encourage them to continue with higher education. Secondly level of education might influence children studying abilities, and finally education is positively correlated with incomes suggesting larger financial endowments of high educated families which can afford to place their children into a track that is more likely to continue with university compared to technical or professional studies.The relationship between school choices and parental occupation is described in Table 2. General high school educated children normally come from families where fathers are specialised white collar (45%) or professional- manager (20%), while nearly 60% of children who attend technical or professional schools have parents working as white collars. At the end teachers' children tend to graduate in general schools (especially if their mother is a teacher 35%). The occupational condition of graduates is influenced by their high school's performance, in fact students with high grades tend to be more selective than the others in the decision of entering labour market or continuing studying and prefer to achieve higher education delaying the entrance in the labour market.

[12] All Tables and Figure are in Appendix A

The distribution of high school grades differs a lot within each type of institute: we notice that students coming from general schools who obtain high grades are nearly 50% while those from technical/vocational schools are the half (25 %, Figure 2). But considering high school grades as indicator of student's capabilities, we are relying on a biased indicator, in a certain sense, because it doesn't really reflect the ability of pupils as they attend different institutes that offer different levels of education (general for liceo versus technical/ professional for istituto tecnico or professionale). Otherwise we consider a good indicator of ability the academic performance previous to high school, given by junior high school's grades. In fact from Figure 3 we derive that more talented pupils tend to go to general schools rather than technical/professional schools, then are more likely to continue with university studies. Looking at this figure we derive an important stylised fact: those students who performed low grades in junior high school are more likely to graduate from professional or technical institutes, while those who obtained high grades (ottimo or distinto) go to general schools. The conclusion is that the better is the performance in junior high school more chances children have to graduates from general schools and consequently to continue studying. Junior high school marks can be considered a measure of ability of children. The choice of continuing studying or going to work is represented in Table 3. The 60 % of leavers go to work after graduating while nearly the 39% continue with studies at the University, of these the 5% interrupt their studies in the first three years of courses. Going back to our measure of ability, junior high school grade, in Table 6 we see that the students who give up with studies in the first three years of university are those who obtained the lowest junior high school's grades, therefore characterised by low ability. Finally Figure 11 presents a decisional flow-diagram for high school leavers. In the figure are indicated two decisional nodes:node (0) where high school leavers decide wherther to continue studying or go to work while node (1) illustrates the decision of continue studying or drop out. In order to shed light on the causal link between high school leavers decisions to continue studying/going to work and subsequent outcomes an empirical investigation using propensity score techniques is carried out.

5. Results

Table 6 presents the impact of returns to higher education estimated by Ols (first row), Fully interacted Ols (second row) and single treatment – matching (remaining rows), while Table 7 present Multiple treatment results. The following discussion examines in depth the results obtained using each method.

a) Ols and Single treatment matching

In this model we estimate returns to undertaking education considering as a treated group potential university graduates, while as a non treated group we consider a group constituted by workers and drop out together then drop out and workers singularly. We begin by performing Ols estimates, then interacted Ols and finally pass to single treatment matching procedure. In the matching procedure we use nearest neighbor matching and kernel matching specifications. Asymptotically, all the different matching estimators produce the same estimate, because in very large sample, they all compare only exact matches. In our case, that of finite samples, different matching estimators produce different estimates because of how much weight they assign to observations and how they handle (implicitly) the common support. Both Nearest Neighbour matching and kernel matching performed very close results. Performing matching treated and untreated individuals have been matched on the basis of the estimated propensity score. The variables included in the probit estimations are all dummies variables: female, if have brothers or sisters, region of residence, type of high school, high school's grade, junior high school's grade, parents education, grandparents education and parents occupational status at 14[13]. The estimated coefficients of all probit model estimated have expected signs and seem to reflect the results derived from the descriptive analysis in all five cases. We found, in fact that parental education, general education and academic performance are all significant and influence positively the probability of continuing studying. General results are presented in table 6. Ols estimated return to higher

[13] Age was not included because we are considering individuals of the same cohort who have almost the same age.

education is around 50% for the first three specifications where we consider as treated group potential undergraduate and as a control group workers, drop out and drop out plus workers respectively. Secondly considering a group of potential university graduates and drop out students as treated units compared to workers we obtain a fairly smaller return around 40%. Finally the comparison between drop out students as the treated group to workers gives a negative effect of nearly 3%. Matching shows that who undertake higher educated enjoys on average a 77% earning premia (ATT) on earnings realised three years after graduation[14], while the estimated return for those who did not achieve any further education would have been around 40% in all the first three specifications. Once again the estimated average return dimishes to 70% comparing the group of potential undergraduates and drop out (treated) with workers (untreated). Finally we find small differences between Nearest Neighboor method and Kernel Matching when we compare drop out students to workers, suggesting high sensitiviity for these comparison groups to weights assigned to observations. All estimates are performed with replacement, allowing many treated units to be matched to the same non treated unit and no exclusion restrictions are included in estimating ATT. On line with literature OLS estimates prove to under-estimate the returns to education for individuals entering the job market.

As seen in the theorical analysis if we allow for heterogeneous returns standars Ols estimates produce bias estimates of ATT. When we use a interacted Ols specification, which allows for all interactions between the Xs and the treatment indicator S, we obtain a result very similar to that obtained with matching. The statistical significance of the interaction terms provides evidence on the presence of heterogeneouse returns to higher education, furthermore both the interacted Ols and single treatment matching estimates of ATT are significantly different from the corresponding ATNT suggesting that if those who did no go on with education had instead undertaken it they would have experienced an higher return than those who are more likely to obtain higher education qualifications. According to the results we see that achieving university education (or at lest attempting to doing so) has a positive effect on earnings, and this effect is much higher when we compare workers and individuals who enrolled at university in 1998.

[14] The third specification is charactterised by the failing of tha balancing property fails probably because of the small number of controls

c) Matching methods: multiple treatment

We now look at a more disaggregated analysis that focuses on the sequential nature of educational levels. We consider a multiple treatments procedure starting with potential university graduates (*PUNG*) versus drop out students (*drop*) and *workers*. As pointed out in Section 3 treatments are sequentially ordered with respect to the level of education attained and the propensity score is estimated using an ordered probit model. The results of multiple treatment are reported in Table 7. Matching shows an average population return of 35% from graduating compared with leaving school immediatly after high school. The comparison between drop out and potential university graduates gives an estimates much higher but not significant, probablly because of the small number of individuals in the control group. In order to examine wherter there is some heterogeneity in the treatment effect between women and men, the sample is divided by sex, and the matching procedure is applied to analyse the average treatment effect by sex.

In brief there is heterogeneity between sexes. Both treatments present higher average treatment effect on the treated (ATT) and average treatment effects (ATE) for men than for women. Consequently achieving higher education appears to be significantly better in terms of earning differential for men. Results remain non significant when comparing drop out students with potential univeristy graduates. While OLS estimation would have hidden such a comparability problem we find that the smaller is the potential comparison group compared to the treated group the harder is to balance their characteristics X_i, revealing that the data simply do not contain enough informations to perform non-parametric estimation.

6. Conclusions

The problem of evaluating the impact of a policy is a central problem in social sciences and this paper presents an empirical investigation of a popular method of evaluating social policies that is matching. A stylised facts of the theory of human capital suggests that if one individual's education is increased of one year he, once entered in the labour market, will earn much more: the aim of this research is to evaluate the returns of high education on earnings using propensity score techniques.

Unlike great part of literature we do not estimate annual wage premia but the effect of higher education on earnings observed three years after graduation. In this paper we analyse the decision nodes indicated in Fig.11., in the first (1) high school graduates decide whether to work or to continue studying, in the second (2) whether to graduate or drop out. These two nodes, in a single treatment model, represents our "alternative treatments" while in a multiple treatment framework represent the sequentional educational levels. We begin our analysis using a OLS specification, then pass to a single treatment framework focusing on the impact of a specific educational level. Then we consider a multiple treatment model that distinguishes between the impact of different educational aquired skills on earnings observed three years after graduation. We allow both for homogeneous and for heterogeneous returns. The parameters of interest defined in tha paper are three: Average treatment effect (ATE), Average Treatment Effect of the treatment on the treated (ATT) and Average Treatment Effect of the treatment on the untreated. In the homogeneous treatment model these parameters are all equal while in the heterogeneous model might differ.The the main goals of this paper can be summarised as follows: first it uses a "new" dataset that includes information on earnings, on educational level and on other relevant individual characteristics that is more informative than the Survey on Income and wealth of the Bank of Italy extensively used by Italian labour economists since now. Second, in line with other evidence we find that OLS under-estimate the marginal return to additional education. In fact, ATT estimates, eliminating the selection bias for P_i in the common support and controlling for ability, give a sizeable and significant return to high educational qualifications. Third, the overall returns to education remain at each level significant [15] even allowing for heterogeneous returns to higher educationan in particular we find an average return of population of 35% for those completing higher education versus high school graduates and an average population return of 67% (non significant) for those who are potential graduates compared to drop out students. Finally we find evidence of heterogeneous return to higher education in terms of observables.

[15] Except in the case of drop out students considered as control group or treated group both in single and multiple treatment.

References

Angrist J. and Krueger, A. B. (1991), 'Does Compulsory School Attendance Affect Schooling and Earnings', Quarterly Journal of Economics, 106, pp.979-1014.

Augurzky B. and Schmidt C. (2001), "The Propensity Score: A mean to an end", IZA Discussion Paper No 271

Becker S. and Ichino A.(forthcoming in the Stata Journal), "Estimation of average treatment effects based on propensity score"

Black D. and Smith J., " How robust is the evidence on the effects of college quality? Evidence from matching" ,WP2003/03

Blundell R., Diaz M.C., Meghir C. and Van Reenen J., "Evaluating the employment impact of a mandatory job search assistance programme" The Institute for Fiscal Studies, London, WP01/20

Blundell, R., Dearden, L., Goodman, A. and Reed, H. (2000), 'The returns to higher education in Britain: evidence from a British cohort', *Economic Journal*, 110, F82–F99.

Blundell, Dearden and Sianesi, (2004) , "Evaluating the impact of education on earnings in the UK : models, methods and resuts from the NCDS, The Insitute for Fiscal Studies, WP03/20

Cappellari L. ," The effect of high school choices on academic performance and early labour market outcomes", Quaderni dell'Istituto di econmia dell'impresa e del lavoro,35/2003

Card, D. (1995), 'Earnings, Schooling, and Ability Revisited', in S. Polacheck (ed.), *Research in Labor Economics*, Vol.14, JAI Press: Greenwich, Connecticut, pp.23-48.

Card, D. (1999), 'Education and Earnings', in O. Ashenfelter and D. Card, editors, *Handbook of Labor Economics*, North Holland: Amsterdam and New York

Checchi D.,” University Education in Italy”, 2000

Chesher A., " Nonparametric identification with discrete endogenous variables"

Dearden, L. (1999a), ‘The effects of families and ability on men’s education and earnings in Britain’, *Labour Economics*, 6, 551–67.

Dearden, L. (1999b), ‘Qualifications and earnings in Britain: how reliable are conventional OLS estimates of the returns to education?’, Institute for Fiscal Studies, WP no. 99/7.

Dehejia, H.R. and Wahba S.(1998), "Causal Effects in non experimental studies: re-evaluating the evaluation of training programmes", NBER working paper 6586

Dehejia H.R. (2001),"Program Evaluation as a Decision Problem" forthcoming in the Journal of Econometrics

Dehejia H.R.(2003), "When is ATE enough? Risk aversion and inequality aversion in evaluating training programs", unpublished paper

Duflo, E. (1999), ‘Schooling and Labor Market Consequences of Scholl Construction in Indonesia: Evidence from an Unusual Policy Experiment’, Mimeo, MIT.

Eichler, M. and Lechner M. (2002), "An evaluation of public employment programmes in the East German State of Sachsen-Anhalt", Labour Economics, 264

Harmon, C. and Walker, I. (1995), ‘Estimates of the Economic Return to Schooling for the United Kingdom’, American Economic Review, 85, pp.1278-1286.

Heckman,J.(1979),"Sample selection Bias as a specification Error", Econometrica,47,153-61

Heckman J. , R. LaLonde and J. Smith (1999), " The Economics and Econometrics of Active Labour Market Programs", Handbook of Labour Economics, vol.III, Ashenfelter and Card Editors.

Heckman J., Ichimura H. and Todd P.(1998), " Matching as an Econometric Evaluation Estimator", Review of Economic Studies ,65: 261-294

Hotz J., Xu L., Tienda M. and Ahituv A.(1999),"Are there Returns to the wages of young men from working while in school?", NBER working paper 7289

Ichino A., Nannicini T. and Mealli F. "Sensitivity of matching estimators to unconfoundedness. An application to the effect of temporary work on future employment", submitted 2004

Imbens G. (2000), " The role of propensity score in estimating dose-response functions", Biometrika, 87,706-10

Imbens G. (2004), "Semiparametric Estimation of Average Treatment Effects under Exogeneity: A Review", Review of Economics and Statistics, 86,4-29

LaLonde, R. (1986), "Evaluating the Econometric Evaluations of TrainingPrograms", American Economic Review, 76,604-620.

Larsson L., "Evaluation of Swedish Youth labour market programmes", WP 2000/06

Lechner M. (1999), "Programme Heterogeneity and Propensity Score Matching: An application to the Evaluation of Active Labour Market Policies",University of St. Gallen, unpublished paper

Lechner M. (2000), " A note on the common support problem in applied evaluation studies", University of San Gallen, Department of Economics, Discussion paper 2001:01

Navarro-Lozano S.,” Matching, Selection and the propensity score:Evidence from training in Mexico”,2003

Origo, F.(2003), "The effects of wage flexibility and the role of institutions: incentive, sorting and rent sharing", Istituto per la Ricerca sociale.

Joffe M. and Rosenbaum P. (1999), "Propensity Scores", American Journal of Epidemiology, 150:327-333

Rosenbaum P. and Rubin D.(1983),"The Central Role of the Propensity in Observational Studies for Causal Effects", Biometrika, vol.70, No 1, pp 41-55.

Rubin, D. and Neal, T. (1996), " Matching using estimated propensity scores: Relating Theory to Practice", Biometrics, 52:249-264.

Rubin, D. (1997), " Estimating Causal Effects from Large Data Sets Using propensity Scores", Annals of Internal Medicine, Vol 127:757-763

Smith J.(2000), "A Critical survey of Empirical Methods for Evaluating Active Labour Market Policies", Zeitschrift fur Volkswirtschaft und Statistik, Vol 136(3),1-22

Smith J.(2000), "Evaluating Active Labour Market Policies: Lessons from North America", unpublished paper

Sianesi B.(?), "An evaluation of the active labour market programmes in Sweden" , The Institute for Fiscal Studies, London, CEMMAP WP 01/05

Sianesi B.(2003), "Differential Effects of Swedish Active Labour Market Programs for Unemployed Adults in the 1990s", (unpublished paper)

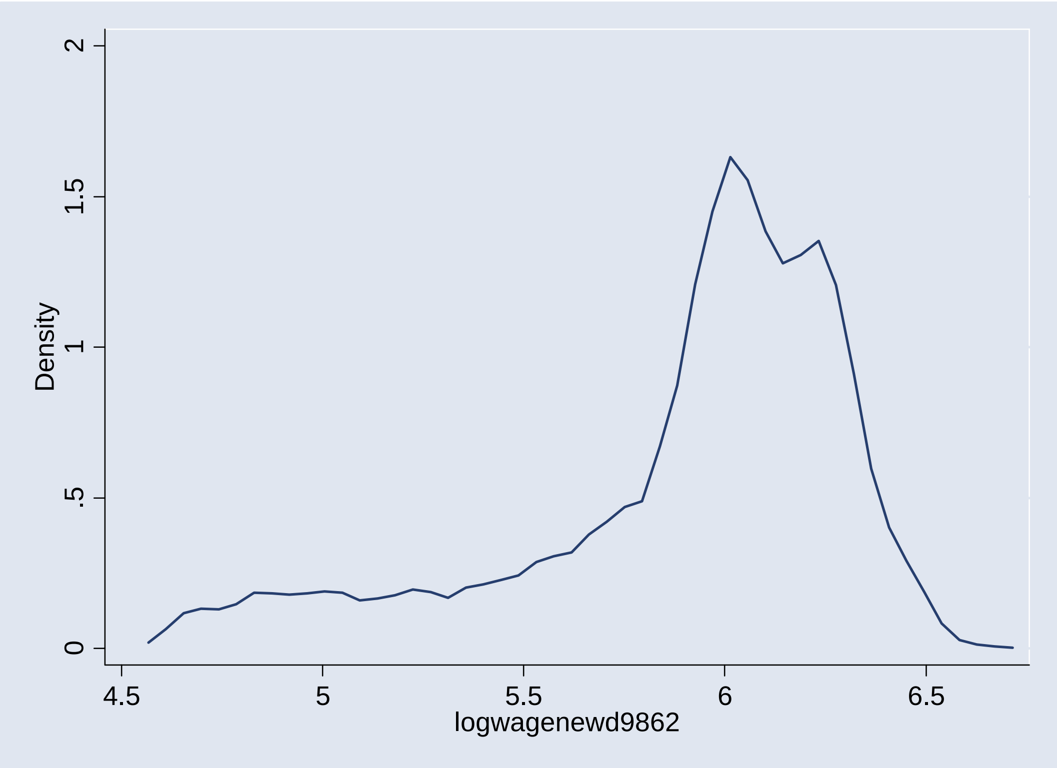
APPENDIX A: Description of the variables used in estimations

- Logwage is the natural logarithm of individuals earnings
- Brosist=1 if the individual has brothers or sisters
- Female=1 if the individual is a female
- Nord=1 if the individual lives in the northern regions of Italy (Piemonte, Lombardia, Liguria , Veneto, Friuli Venezia Giulia,Trentino Alto Adige)
- Centro if the individuals lives in the centre of Italy (Emilia Romagna,Umbria, Lazio, Toscana)
- Isttec=1 if the individual attended a technical high school
- Istprof=1 if the individual attended a vocational high school
- Liceo=1 if the individual attended a general high school
- Genistruito=1 if at least one of individual parents' achieved high school education or higher education
- dvmat_42_47=1 if individual high school grade is the range 42-47
- dvmat_48_54=1 if individual high school grade is the range 47-54
- dvmat_54_60=1 if individual high school grade is the range 54-60
- dvmedbuod01=1 if individual junior high school grade is buono
- dvmeddstnd01=1 if individual junior high school grade is distinto
- dvmedottid01 if individual junior high school grade is ottimo
- fatheremploy=1 if individual's father was employed when individual was 14 years old
- motheremploy=1 if individual's mother was employed when individual was 14 years old
- grandparenteduc=1 if at least one of grandparents received high school education or higher

Appendix B: The dependent variable (log earnings)

Our dependent variable is log earnings. The questionnaire contains only banded earnings data, furthermore data are topcoded so that we only have a lower bound for each individual. Therefore we re-built predicted earnings for each individual using interval regression.[16]

In Figure A1.1 is represented the distribution of log earning of high school leavers.



[16] We use the INTREG procedure in Stata 8 which is a robust estimation procedure that takes into account for the complex sample structure when calculating point estimates and standard errors. See Forth and Millward (2000) for the log likelihood function and details of the estimation methodology.

Moreover to construct a continue variable of earnings for students who dropped out university before completing a degree and of students enrolled at university we use a two stage approach: first we estimate earnings university graduates' earnings using data on from the "Survey on the transition from university to work for university graduates of 1998"[17], and second we predict earnings using the estimated β's and combining with the characteristics of drop out and students enrolled in university.

[17] Once we eliminated observations for graduates in medicine who are likely to get out of average earnings in the specialisation period.

APPENDIX C: TABLES AND FIGURES

Tab.1:High school degree types and parental education

	General education		Technical / professional education	
	father education	mother education	father education	mother education
none or elementar	8.5	11.2	25.7	29.2
Junior high school	25.7	31.3	46.0	43.8
High school	43.7	48.9	25.7	22.4
University	22.4	8.6	2.5	6.2

Tab. 2:High school degree types and parental education

	general education		Technical / professional education	
	father education	mother education	father education	mother education
self employed	8.1	6.5	11.0	12.2
Craft	4.9	1.4	8.0	2.9
Entrepreneur	5.6	1.4	4.4	1.3
Professionnel	13.4	3.6	5.8	2.1
Manager	11.3	2.9	1.6	0.5
White collar high level	44.9	41.4	52.6	48.5
White collar low level	6.2	7.0	15.2	21.7
Teacher	5.6	35.8	1.4	10.8

Table 3: Decision of high school leavers

	N	%
a)Go to work	14015	60.2
b)Interrupt studies	1274	5.4
c)continue studying after 3 years	7973	34.4
(b+c)Enrol at University	9247	39.6
Total	23262	100

Tab 4: Summary Statistics

	Individuals enrolled at University in 1998	Students who drop out from University	Students who in 2001 are still enrolled at University	Individuals who enter the job market	Total sample
Personal characteristics					
Age	2.17	2.38	2.14	2.5	2.4
nationality	0.99	0.99	0.99	0.99	0.99
Sex	0.61	0.53	0.61	0.5	0.51
brothers/sisters	0.83	0.84	0.83	0.88	0.86
zone of residence					
North	0.36	0.33	0.36	0.38	0.43
South	0.35	0.34	0.33	0.33	0.29
Center	0.33	0.33	0.33	0.31	0.33
Academic Curricula					
Junior high school	2.49	2.07	2.55	1.6	1.79
high school	2.51	2.16	2.5	1.8	2.1
type of high school					
Liceo	0.4	0.17	0.43	0.05	0.11
Technical studies	0.34	0.44	0.33	0.36	0.36
professional studies	0.17	0.31	0.15	0.53	0.46
Other studies	0.07	0.07	0.07	0.07	0.06
Parental Occupation					
Mother white collar	0.4	0.34	0.41	0.27	0.3
Father white collar	0.68	0.67	0.68	0.69	0.7
Mother self employed	0.07	0.08	0.07	0.07	0.07
Father self employed	0.29	0.28	0.29	0.27	0.29
Parental Education(if high)					
Mother	0.44	0.34	0.45	0.17	0.22
Father	0.47	0.37	0.48	0.19	0.25
granparents	0.15	0.11	0.16	0.05	0.06
Nobsv	**9247**	**1274**	**7973**	**14015**	**23262**

Table.5: Probit estimates

	(1)		(2)		(3)		(4)		(5)	
variable	***coef***	***z***	***coef***	***z***	***coef***	***z***	***coef***	***z***	***coef***	***z***
female	0.018	3.09	0.068	3.29	0.10	2.86	0.05	2.34	-0.04	-1.20
	0.02		0.02		0.35		0.03		0.03	
brosist	-0.13	-4.55	-0.129	-4.43	0.01	0.31	-0.15	-5.15	-0.12	-2.75
	0.03		0.03		0.05		0.28		0.04	
nord	-0.22	-9.67	-0.2	-9.26	0.02	0.51	-0.14	-6.49	-0.14	-3.99
	0.02		0.02		0.04		0.021		0.03	
centro	-0.08	-3.42	-0.07	-3.25	-0.03	-0.70	-0.09	-0.45	0.02	0.46
	0.02		0.02		0.04		0.021		0.03	
isttec	-0.25	-6.41	-0.25	-6.76	-0.19	-2.76	-0.15	-4.09	0.04	0.06
	0.04		0.04		0.07		0.036		0.06	
istprof	-0.75	-18.69	-0.72	-18.65	-0.41	-5.76	-0.61	-16.33	-0.23	-3.68
	0.04		0.04		0.07		0.037		0.06	
liceo	1.1	23.72	1.03	23.37	0.36	5.01	0.89	21.19	0.52	6.92
	0.05		0.04		0.07		0.042		0.07	
genistruito	0.44	19.6	0.42	19.42	0.06	1.54	0.44	21.11	0.35	10.35
	0.02		0.02		0.04		0.21		0.03	
dvmat_42_47	0.34	13.37	0.34	13.59	0.029	. 4.62	0.30	12.87	0.16	4.23
	0.03		0.02		0.45		0.023		0.03	
dvmat_48_53	0.62	20.58	0.59	20.56	0.03	6.51	0.55	19.82	0.28	6.28
	0.03		0.04		0.50		0.027		0.44	
dvmat_54_60	1.02	29.05	0.97	28.89	0.52	9.37	0.92	28.20	0.46	8.29
	0.03		0.03		0.06		0.03		0.55	
dvmedbuod01	0.1	4.02	0.09	3.71	0.01	-0.17	0.10	4.26	0.08	2.15
	0.03		0.03		0.43		0.023		0.03	
dvmeddstnd01	0.3	9	0.27	8.67	0.07	1.33	0.29	9.49	0.20	4.08
	0.03		0.04		0.52		0.03		0.05	
dvmedottid01	0.42	10.23	0.41	10.73	0.20	3.25	0.42	11.03	0.21	3.08
	0.04		0.04		0.06		0.037		0.06	
fatheremploy	*0.04*	*0.86*	*0.11*	*3.55*	-0.0008	-0.02	0.03	0.70	0.0009	0.01
	0.05		*0.04*		0.05		0.043		0.06	
motheremploy	*0.13*	*6.02*	*0.35*	*1.04*	0.004	0.07	0.11	5.49	0.078	2.38
	0.02		*0.06*		0.06		0.2		0.03	
nonnistruiti	*0.25*	*6.44*	*0.233*	*6.31*	0.09	1.64	0.26	7.10	0.175	2.94
	0.04		*0.06*		0.06		0.036		0.05	
_cons	*-0.75*	*-11.17*	*-0.78*	*-13.19*	0.74	7.12	-0.74	-11.74	-1.54	-15.08
	0.07		0.06		0.10		0.062		0.10	
nobsv	21988		23262		7973		23262		15289	

(1)The treated group is composed by potential university graduates while the control group by workers; (2) the treated group is formed by potential university graduates and the controls by drop outs plus workers; (3) the treated group is composed by potential university graduates and the treated group by dop out students; (4)the treated group is composed by potential university graduates plus drop out while the treated group by workers and finally (5) the treated group is formed by drop out and the controls by workers.

Tab. 6: The returns to higher education compared with less than higher education (Single treatment, % earning gain)

Treatment	N(1)	comparison	N(0)	ATT**	ATE	ATNT
			OLS (ATT=ATE=ATNT)			
PUNG	7973	*workers*	14015		0.53	
PUNG	7973	*drop+workers*	15289		0.54	
PUNG	7973	*drop*	1274		0.56	
PUNG+drop	9247	*workers*	14015		0.4	
drop	1274	*workers*	14015		-0.029	
			FULLY Iteracted OLS			
PUNG	7973	*workers*	14015	0.8	0.52	0.37
PUNG	7973	*drop+workers*	15289	0.8	0.53	0.38
PUNG	7973	*drop*	1274	0.8	0.7	0.54
PUNG+drop	9247	*workers*	14015	0.69	0.44	0.27
drop	1274	*workers*	14015	-0.029	-0.017	-0.016
			Nearest Neighbor Matching			
PUNG	7973	*workers*	14015	0.79	0.53	0.37
PUNG	7973	*drop+workers*	15289	0.798	0.52	0.38
*PUNG**	7973	*drop*	1274	0.77	0.73	0.54
PUNG+drop	9247	*workers*	14015	0.688	0.424	0.253
drop	1274	*workers*	14015	-0.027	-0.047	-0.045
			Kernel Matching			
PUNG	7973	*workers*	14015	0.76	0.51	0.36
PUNG	7973	*drop+workers*	15289	0.77	0.51	0.385
*PUNG**	7973	*drop*	1274	0.73	0.55	0.71
PUNG+drop	9247	*workers*	14015	0.66	0.43	0.28
drop	1274	*workers*	14015	-0.043	-0.052	-0.053

Notes to table 6.:

*The balancing property is not satisfied.Controls used in each specification:region, standard family background information, high school variables, high school's scores, junior high school's scores.

**Average effect of the treatment on the treated (ATT), Average treatment effect (ATE), Average effect of the treatment on the untreated (ATNT)

Table 7: Treatment effects by higher qualification achieved (Multiple treatment approach, % earning gain)

treatment	N_1	comparison	N_0	ATT	ATE	Median Bias Before[1]	Median Bias After	p>chi2[2]
>3 years University	7973	*none*	14015	0.46	0.35	24.7	0.98	0.225
		<3years	1274	0.69	0.67	13.54	2.38	0.019

A)MALE

treatment	N_1	comparison	N_0	ATT	ATE	Median Bias Before	Median Bias After	p>chi2
>3 years University	3100	*none*	613	0.64	0.47	30.98	1.288	0.485
		<3years	6974	0.76	0.75	14.95	3.74	0

B)FEMALE

treatment	N_1	comparison	N_0	ATT	ATE	Median Bias Before	Median Bias After	p>chi2
>3 years University	4873	*none*	7041	0.35	0.34	25.27	1.34	0.187
		<3years	661	0.54	0.52	11.4	2.94	0

Notes to Table 7:

(1)Following Rosembaum and Rubiun (1985) the median absolute bias before and after matching is calculated. The standardised difference *before* matching is the difference of the sample means in the full treated and nontreated subsamples as a percentage of the square root of the average of the sample variances in the full treated and non treated groups.The standardised bias *after* matching is the difference of the sample means in the macthed treated and matched non treated subsamples asa percentage of the square root of the average of the sample variances in the full treated and non treated groups.

$$B_{before} = 100 \frac{X_1 - Xx_0}{\sqrt{(V_1(X) + V_2(X)/2}} \qquad B_{after} = 100 \frac{\overline{X}_1 - \overline{X}_0}{\sqrt{(V_1(X) - V_2(X)/2}}$$

(2) p-value of the likelihood-ratio test after matching testing the hypothesis that th regressiors are jointly insignificant,i.e well balanced in the two matched groups

Fig 1: Type of institute attended by high school graduates of 1998

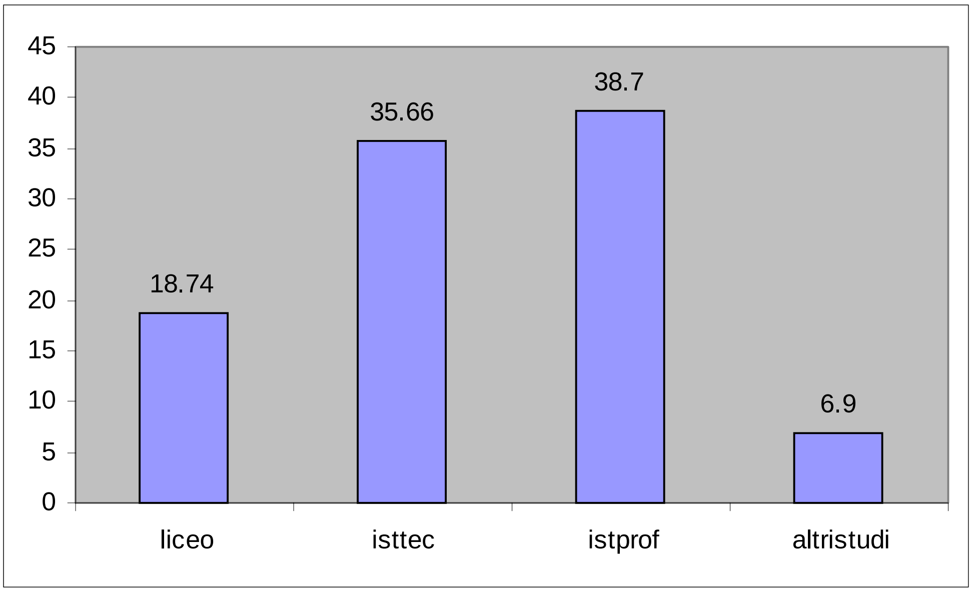


Fig.2:High school grades and high school types

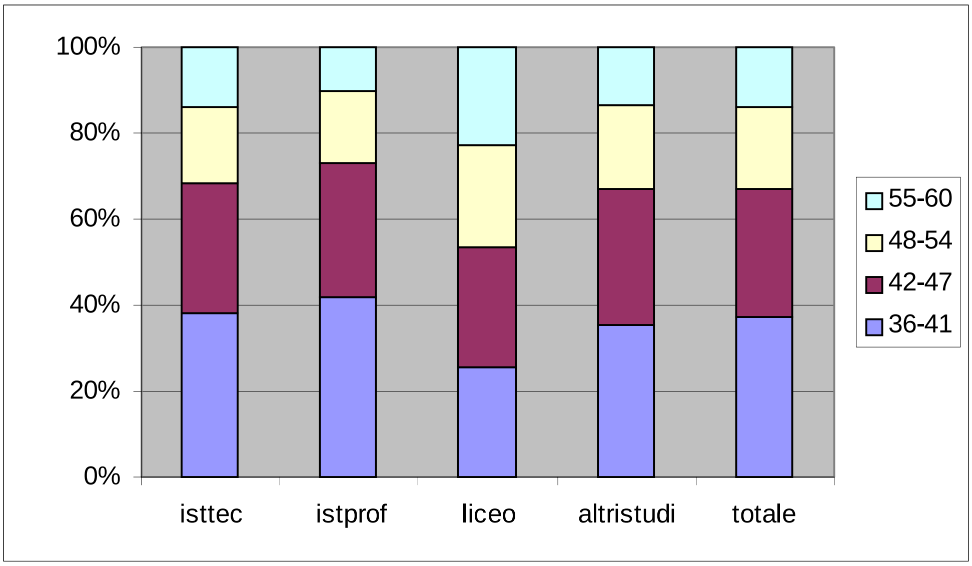


Fig.3: Junior high school grades and high school types

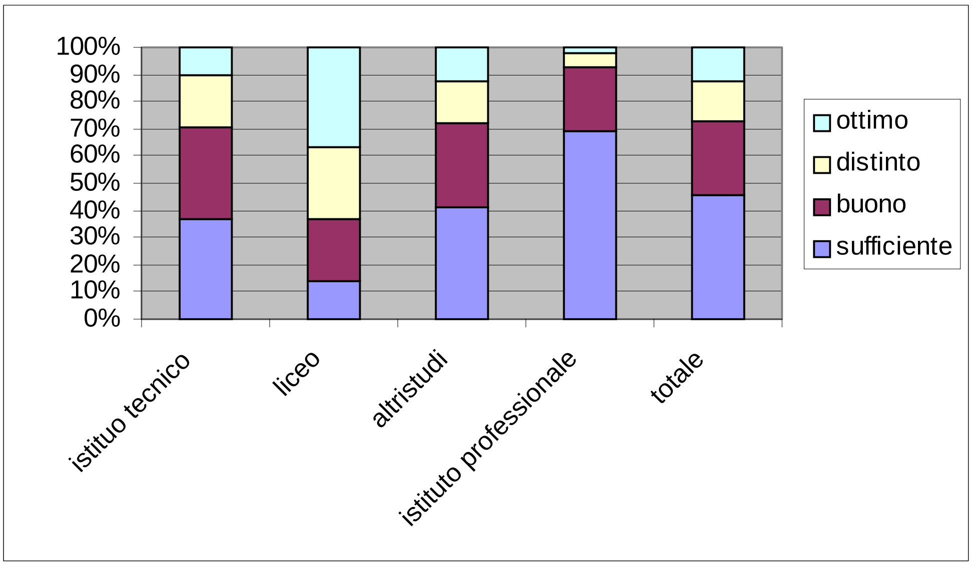


Fig.4:Partecipation in further studies by type of high school in 1998

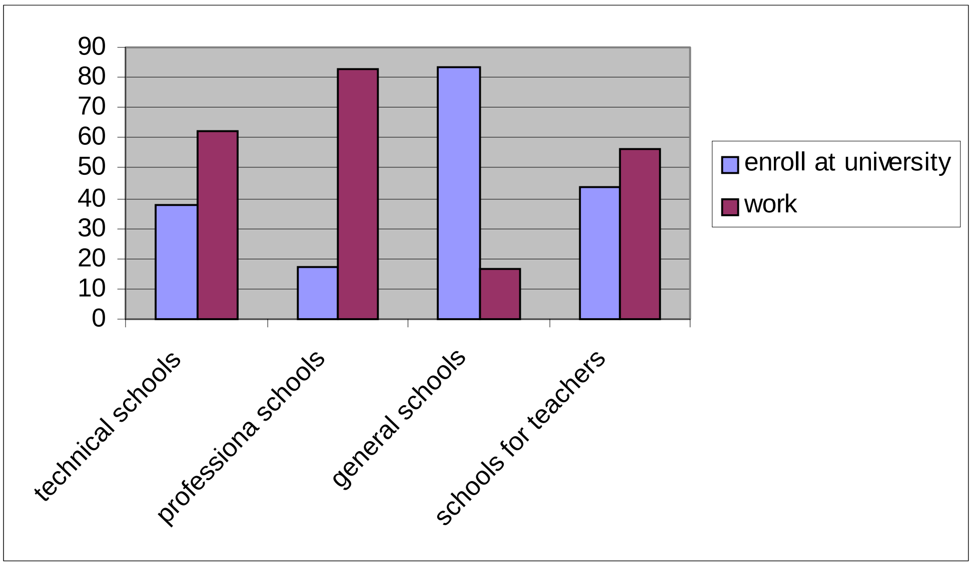


Fig. 5: Type of degree attended at university

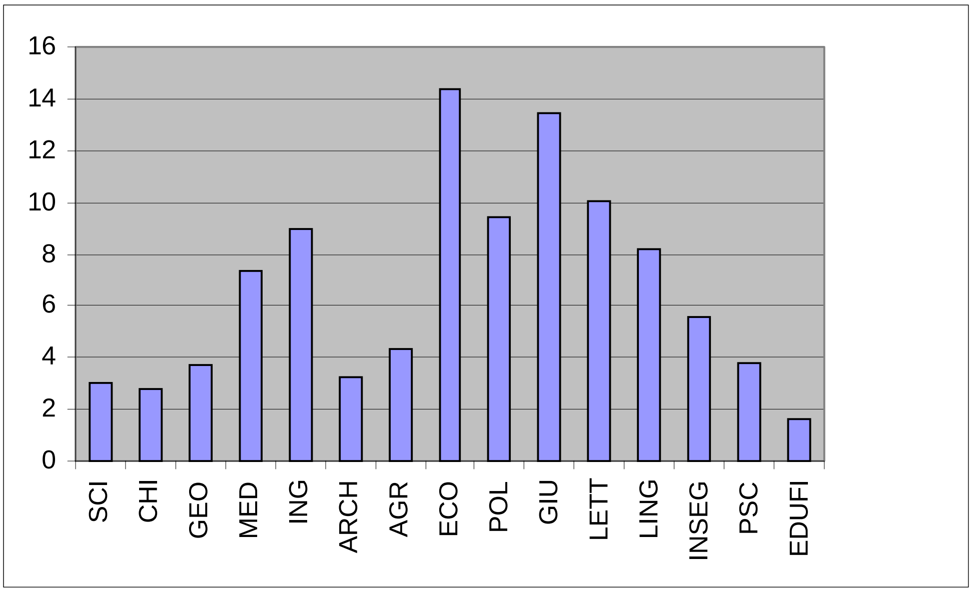


Fig.6: Junior high school grades of drop out and university students

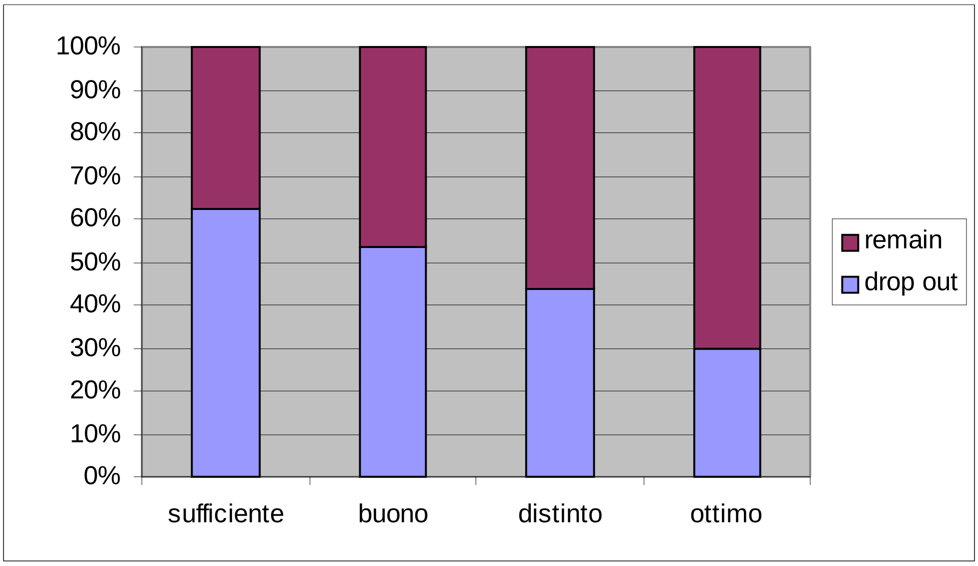


Fig 11: Flow-diagram of the decision of working or studying for high school leavers: