VERY PRELIMINARY DRAFT

HIGH SCHOOLS AND LABOUR MARKET OUTCOMES: ITALIAN GRADUATES, 1995.

By

Dario Pozzoli[1]

Graduate School in the Economics and Finance of Public Administration

Catholic University, Milan

July, 2006

Abstract

The main objective of this paper is to provide empirical evidence on differences in early occupational labour market outcomes across high school tracks. I consider a multiple treatment model, which distinguishes the impact of four different types of high school, thus allowing the attainment of different educational qualifications to have separate effects. The paper estimates relative log-wage premium of employed graduates by high school type, three years after graduation. I used a propensity score matching-average treatment on the treated method to correct directly both for student self-selection into school type and for the decision to participate to the labour market. Using a large data set from a survey on job opportunities for the 1995 Italian high school graduates conducted by the Italian National Statistical Institute in 1998. I found that Professional and Vocational high school graduates have generally a positive wage premium with respect graduates from General and Teaching/art high school types. These effects seem to be greater for women than for men. Moreover the estimation of the model where participation and employment are considered as post-treatment effects of the high school type suggest that vocational and technical education increases not only early earnings but also participation to the labour market and employment probability.

Keywords: matching estimator, multiple treatment, returns to education, selection bias.

JEL classification: J64, C41, C50.

[1] E-Mail: dario.pozzoli@unicatt.it

1. Introduction

This paper is concerned with those youths who have completed secondary school education. This study, in particular, examines the efficacy (in terms of labour market outcomes) of the vocational/technical education in Italy as compared with that of academic schools, using a large data set from a survey on job opportunities for the 1995 Italian high school graduates conducted by the Italian National Statistical Institute in 1998.

The Italian secondary school system consists of four types of high schools: licei (General high schools), istituti tecnici (Professional high schools), istituti professionali (Vocational high schools) and finally Teaching and Art schools. The first type of school last five years and provide their graduates with general and academic skills useful for higher education. The second and the third type are generally five-year and provide their students with professional skills that can be exploited in the labour market immediately after graduation. However graduates from either Professional or Vocational high school are allowed to enrol at university if they choose to go on to further education. The fourth type- typically for those students who desire to become teachers- last four years plus an additional year for those students who wish to go to University. So students from any of these track are formally entitled to enrol at University, conditional on having attended 5 years of secondary school. However, as documented by Cappellari (2004), each of these tracks predicts very different outcomes in terms of additional schooling acquired: for example, graduating from General high schools, compared to Technical/Vocational ones, substantially increases (+60%) the probability that individuals go on to university rather than becoming active on the economic front. This documents that early educational career decisions have a strong effect on the choices available at later educational stages.

In Italy, the choice on the type of the secondary school to be attended is typically taken at the age of 13, during the final year of junior high school. So we can consider the Italian system as an early decision track system: this implies a trade-off between specialization (efficiency) and misallocation (equality). On one hand, early differentiation can improve teacher effectiveness because teachers can target instruction at the level more closely aligned with student needs than is possible in more heterogeneous environments (Figlio et al 2000). On the other, the earlier the tracking occurs, the most likely it is that selection of pupils into tracks is related to parental background (family wealth, parental education) rather than to student ability and the easier it is to decrease intergenerational mobility and equality. Several papers stressed and documented the costs of misallocating pupils implied by the tracked high school system when sorting is not based on cognitive ability but on parental background (Dustmann 2004; Hanushek and Woβssman 2005; Brunello et al 2005; Checchi and Flabbi 2006).

Nevertheless, in this paper I will focus only on the positive side of the Italian upper secondary school, i.e. the gains from specialization implied by a stratified system. To prove the efficiency of a system organized into tracks, I will provide empirical evidence on differences in early occupational labour market outcomes between graduates in General (Licei), Vocational (Istituti Professionali), Professional (Istituti Tecnici) and Teaching/art schools (Istituti d'arte/Istituto magistrale). The analysis is carried out on a relatively short post-education period, i.e. three years after graduation. I think, however, that this will not constitute a problem for three main reasons: (i) early labour market outcomes are generally good predictor of the subsequent lifetime outcomes; (ii) focusing the analysis on a short period after graduation will make counterfactual analysis easier, because in this way I will only compare individuals with the same level of education excluding the higher ones; (iii) variation in entry wage of high school graduates will mainly depend to variation in educational background and not to their labour market experience.

I will consider a multiple treatment model, which distinguishes the impact of four different types of high school, thus allowing the attainment of different educational qualifications to have separate effects. This paper, in particular, estimates relative log-wage premia of employed graduates by high school type three years after graduation. I use a propensity score matching-average treatment on the treated method (PMS-ATT model) devoting great attention to endogenous selection issues in order to unravel the casual link between school types and subsequent outcomes. It is important to note that I uniquely focus on the private return to secondary education three years after graduation, ignoring any potential externalities that may benefit the economy at large. In addition, the average individual return to high school I report is only one component in a full analysis of the private returns to education, which would have to balance individual cost against a flow of such returns over the working life. Moreover, I say nothing about the riskiness of education returns, an important determinant of educational choices among less wealthy families.

Taking into account the endogeneity of both high school choice and labour market participation decision, I find out that both Vocational and Technical high school graduates have a positive wage premium with respect to graduates from General and Teaching/art high school types. These effects seem to be greater for women than for men. Moreover the estimation of the model where participation and employment are considered as post-treatment effects of the high school type suggest that vocational and technical education increases not only early earnings but also participation to the labour market and employment probability.

Hence it seems that for those youths proceeding to the labour market after leaving high school, technical or vocational education is superior to general one in terms of early labour market outcomes three years after graduation. This is in contrast to the world bank's orthodoxy that the skills taught in most vocational and technical tracks are of little value to employers and employees since less flexible and transferable than general skills in a globalized market where non hierarchical firms demand more general and versatile skills (Aghion et al 1999). The results presented above are instead consistent with the supporters of vocational and technical education, for whom this type of education teaches students marketable skills and attitudes that can help them find skilled jobs and reduce their risk of unemployment or employment as low-paid unskilled workers (Collins 1975; Bishop and Kang 1989; Arum and Shavit 1995).

These findings could suggest that all the recent attempts of reforms[2] trying to empty technical education of all meaning were not the best way to renew the secondary school system.

This study has six parts and has the following structure. Section 2 reviews the empirical literature about the effects on early earnings of different high school types. Section 3 describes the theoretical framework. Section 4 is devoted to the description of the data and sample used in the empirical exercise carried out in this study. Section 5 gives an account of the econometric specifications and methods of estimation. Section 6 discusses the estimation results obtained and the final section concludes the paper.

2. Literature Review : Rate of returns of different high school types

The evaluation of the different effects on early earnings of General, Professional and Vocational schools has not been extensively researched. The prevailing orthodoxy, in the literature, is that the rate of return is much higher to investments in general than in vocational secondary education and the accumulated evidence from international case study literature argues that vocational education is best delivered to workers once in employment by enterprises themselves (that is, on-the-job training) with private sector

[2] See in the appendix for more details about the proposed reforms.

training institutions taking the lead in providing formal, off-the-job training where it is necessary.

For Italy, Cappellari (2004) found that general high schools decrease labour market participation and employability while professional schools better the quality of the school-to-work transition, both in terms of participation and employment probabilities. Checchi (2000) finds that attending a professional or a vocational school yields a yearly rate of return of above 6%, whereas attending a general school generates a lower return of 5% for every additional year of education achieved.

Evidence for France shows that professional high school graduates outperform general high school ones in in terms of the time it takes to find their first stable jobs and their earnings once the school to work transition is well established, thanks to the more effective labour market networks they can access (see Margolis and Simonnet 2003). Zymelman (1976) is unable to arrive at any firm conclusion concerning the relative efficiency and effectiveness of general and vocational secondary schooling. More specifically, his review of the five relevant life-cycle rates of return studies in the United States concluded that their findings were contradictory. While life-cycle rate of returns to vocational secondary education were higher in two of the studies, two others reached opposite results, and the remaining study found no difference in life-cycle rate of returns. Psacharopoulos (1987) was able on the basis of evidence from just seven countries (Colombia, Cyprus, France, Indonesia, Liberia, Taiwan and Tanzania) to reach the unambiguous conclusion that the life-cycle rate of return to investments in general curricula is much higher than in vocational/Professional programs. Psacharopoulos (1994) reviewed studies where life-cycle returns to general secondary school are higher than to the vocational track and stressed that the differences in social rates of return is more dramatic because of the much higher unit cost of vocational education. Neuman and Ziderman (1991), using the subset of individuals who were between the ages of 25 and 49 drown from the 1983 Israeli Census of Population and Housing, found that vocational high school graduates who in training related occupations earn more than general high school graduates while there is no significant difference in earnings of vocational school graduates in occupations unrelated to the course of study and those of general school graduates. In Brazil, Arrigada and Zinderman (1992) found that the life-cycle earnings of vocational school graduates in training-related occupations were 16-28 per cent higher than those of academic school leavers, and in Hong Kong, Chung also concludes that the “users” of the vocational and technical education have higher earnings than the general education group. For Turkey, Tansel (1994) found that, vocational/Professional high schools are better than general high schools in terms of labour market outcomes such as life-cycle private rates of return to schooling, unemployment rates and wages. Bennell (1996) critically examined the studies on relative social profitability of general and vocational/professional schooling in developing countries and concluded that no convincing evidence exists that supports the orthodoxy (that has been largely initiated and sustained by World Bank economists) that the social rate of returns to vocational secondary education are generally lower than those to general secondary education. He stressed in fact that among the 19 country studies relying upon data of reasonable quality, only five of them arrived at rate of returns to general secondary education that are significantly higher than to vocational secondary schooling.

3. Theoretical framework for the high school choice

The model I present below is based on school choice model as presented by Todd and Wolpin (2003) and Bratti (2004). I believe that this model can give some useful suggestions for the specification of the econometric model regarding the estimation of the propensity scores to enrol at the different high school types.

Let us assume that an individual has to decide the type of secondary education j, i.e. the high school type, among a set of J available alternatives. For the sake of simplicity I use here, like in the bulk of the literature, linear functional forms. Let the utility of a student depend on her educational performance, her ability endowment and a stochastic term in the following way:

$$U_j = T_j + \gamma_j \mu_0 + \varepsilon_j \quad (3.1)$$

where I have omitted for simplicity the subscript for the individual[3], T_j is performance in high school j, μ_0 is student's ability and ε_j a taste shifter, i.e. an idiosyncratic stochastic term affecting the utility of subject j and unobservable to the econometrician. I have chosen here to let utility depend on educational performance and ability only[4]. However, the argument remains the same if we assume that utility depends on income and that the

[3] The weight given to T_j is assumed to be constant across alternatives and has been normalized to one for simplicity.
[4] I use a non stochastic specification for the utility function. However, the term involving μ0 is usually not observed and enters the error term in empirical applications.

latter is a function of educational performance, or that utility depends on both income and performance, although the notation becomes more involved.

I assume an educational production function of the following form:

$$T_j = \tau_{0j} + \tau_{1j} T_{s-1} + \tau_{2j} F_j + \tau_{3j} \mu_0 \qquad (3.2)$$

where T_{s-1} is the previous educational stage performance, and F_j some family (or educational institutions) inputs. Here, I assume a 'value added approach' also for the 'true' educational production function. I posit that there is a technology (EPF) which transforms educational inputs into an output represented by educational performance or 'knowledge'. In this sense, past knowledge (i.e. performance) is combined with current inputs in order to obtain new knowledge, which is measured by T_j . The crucial assumption is that only knowledge acquired in the immediately previous educational stage and measured by the relative educational performance is used to produce new knowledge (or performance) at each new stage, which is tantamount to assuming that the amount of knowledge acquired by individuals at stages 1,....,s – 2 is embedded in the knowledge acquired at stage s – 1, where s is the current educational stage. Here, I consider two stages only: junior high school and secondary high school. To make an example, in my case I assume that only the knowledge acquired through junior high school is useful for degree performance while primary school performance, for instance, does not give any additional benefit over and above junior high school. In this respect, I assume an EPF different from Todd and Wolpin (2003) who not only assume that the 'true' EPF depends at each stage on the complete flow of inputs up to the educational stage under study, which is also true in my specification but also that past inputs contribute to the production of current cognitive achievement over and above the effect acting through past achievement. To go back to my previous example, in Todd and Wolpin's specification past educational inputs, such as the primary school attended or the resources devoted to education by a student's family when she was a child, have a direct effect on current degree performance over and above that exerted through junior high school. Thus, in my approach, unlike Todd and Wolpin (2003) and like Bratti (2004), T_{s-1} is a sufficient statistic for all educational inputs used in previous educational stages[5].

[5] It is difficult to say which one of the two different views of the cognitive achievement process is closer to reality since little is known about the process through which 'knowledge' is formed. In the future, a major

Plugging equation (3.2) into (3.1) we obtain:

$$U_j = \tau_{0j} + \tau_{1j} T_{s-1} + \tau_{2j} F_j + (\tau_{3j} + \gamma_j)\mu_0 + \varepsilon_j \qquad (3.3)$$

Unfortunately, this specification still includes some variables unobservable to the econometrician. Typically the family inputs F_j are missing in the administrative data commonly used to estimate EPFs. However, I can suppose that family inputs are in turn the outcome of an optimizing process and a function of both observable and unobserved (i.e. missing) family exogenous characteristics. In particular:

$$F_j = \phi_{0j} + \phi_{1j} C + \phi_{2j} M + \phi_{3j} \mu_0 \qquad (3.4)$$

where C are observed family characteristics, such as parents' education and occupation, while M are family unobserved characteristics[6]. By plugging equation (3.4) into (3.3) we obtain the following expression for a student's utility:

$$U_j = \tau_{0j} + \tau_{2j}\phi_{0j} + \tau_{1j} T_{s-1} + \tau_{2j}\phi_{1j} C + \tau_{2j}\phi_{2j} M + (\tau_{2j}\phi_{3j} + \tau_{3j} + \gamma_j)\mu_0 + \varepsilon_j \quad (3.5)$$

Let us define , the error component $\tau_{2j}\phi_{2j} M + (\tau_{2j}\phi_{3j} + \tau_{3j} + \gamma_j)\mu_0 = v_j$, due to potentially observable but missing variables. This gives me some useful insights into possible causes of bias. First, as observed by Todd and Wolpin (2003) conditioning on past education performance (T_{s-1}) makes the model susceptible to endogeneity bias. The endogeneity is due to the correlation between T_{s-1} and the error component v_j . This correlation arises both directly from μ_0, the individual unobserved ability endowment, and indirectly through past family (or school) inputs which both enter T_{s-1} and are likely to be correlated with current family inputs (through μ_0 and M). All these reasons explain why the estimate of τ_{1j} is likely to be biased. However, the bias may also extend to the estimate of $\tau_{2j}\phi_{1j}$, the effect of observed family characteristics. This may happen through the correlation between M and C, i.e. between observed and missing family characteristics, or that between C and

interaction between educational economists, psychologists and educational researchers could give useful insights for a correct specification of EPFs.

6 For simplicity we consider here only time-invariant family characteristics, at least in the period between enrolment in junior high school and the choice of the high school type.

μ_0, i.e between observed family characteristics and student's unobserved ability. In our case, among family observed characteristics the primary focus is on parents' education and occupation, which capture parents' social class.
Thus, the effect of parents' education and occupation is likely to be biased if other unobserved family characteristics are correlated with the former and affect students' school type choice or if students' unobserved cognitive ability endowment is correlated with parents' education and occupation. However, on the grounds that junior high school grades also act as a good proxy for contemporaneous unobserved family inputs and students' ability, the 'value added specification' may help to attenuate the bias in the estimation of the social class effects on degree subject choice in the spirit of the 'proxying and matching' method in Blundell et al. (2000). Indeed, if individuals with different social class origins who enrol in different high school types systematically have different levels of ability the effect of social class might reflect the effect of the latter factor.

4. Data, Sample Selection Criteria and Variables

The analysis is accomplished by using a large data set (18,843 individuals) from a survey on job opportunities for the 1995 Italian high school graduates conducted by the Italian National Statistical Institute in 1998. The sample represents approximately 5 percent of the population of Italian high school graduates of 1995 and contains a wide range of information on the high school curriculum and on post high school labour market experiences. In addition, information on personal characteristics and family background is available. For the present analysis, the sample of 18,843 records has been reduced by eliminating those who were employed and started their job while at high school, since their post-graduation experiences might not be comparable with those of the rest of the sample. The resulting sample size is nearly 18,000 high school leavers. Note that in the empirical analysis, I will take into account both high school leavers who go directly to work and those who enrol at university after graduation.
The main outcome variable is log net monthly earning which is available for only 5980 individuals[7]. Its distribution is given in figure 1 where we can see that it is slightly left skewed with a mean near to 7 and a variance of 0.16.
The definition of explanatory variables used to calculate the wage premia are reported in Appendix 1 along with their sample means. Concerning the covariates the following

[7] I dropped from the original sample the extreme observations (those lower than 1^{th} percentile of the earnings distribution and those higher than 99^{th} percentile).

clarifications should be made: (i) The Italian High Schools have been classified into 4 main categories: General (Licei), Professional (Istituti tecnici), Vocational (Istituti professionali) and Teaching/art. From table 1 we can see that graduates from General high school represent nearly 30.86% of the whole sample, while graduates from Professional and Vocational institutes constitute 44.57% and 14.30% respectively. Finally those who exit from Teaching/art high schools consist in only 10.28%. Males and female are almost equally distributed in all high school types, with the exception of Teaching/art high schools (only 11% of graduates are male); Table 2 report the distribution of high school type within each macro Italian region (north-east, north-west, centre, south): most of students living in the south are enrolled at general high school (nearly 50%) while students in the north of Italy generally prefer technical and vocational education to the general one (only 15% of individuals in the north are enrolled at general high school). (iii) As indicator of student's performance and capabilities I used the variable "high school final mark". The distribution of high school grades differs a lot within each type of school (table 3): students coming from General high school who obtain high scores are almost 21.47% while those from Vocational schools are 11.21%. This indicator could be a biased one because it may not reflect the ability of pupils as they attend different types of education (General versus Vocational or Professional for example). To compensate partially for this, I used as a good proxy of ability also the academic performance previous to high school, given by junior high school's grades. This variable is strongly correlated with the choice of the school type: table 4 shows that nearly 68% of students with high grades in junior high schools go to General schools and that those students who performed low grades instead are more likely to graduate from Professional or Vocational institutes. (iv) Parental background is described by 7 categorical variables summarizing both parents' educational level. There is a positive relation between parents' education levels and the probability that their children get general education. There can be three possible explanations to this situation according to the main literature in this field (Figlio 2000; Checchi and Flabbi 2005; Dustmann 2004; Cappellari 2004; Woessman and Hanushek 2005). 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 track that is more likely to continue with university compared to vocational or professional studies. The relationship between school choices and parental education is described in table 5. General high school educated children normally come from families where both parents have at least high school degree (54%), while nearly 72% (58%) of children who attend Vocational (Professional) schools have both parents with at most junior high school degree.

5. Methodology

This section describes two alternatives models used to estimate graduate log-wage premia by school type.

5.1 Linear regression (OLS)

The regression model can be described as follows. Let us assume that the post-high school earnings of individual i, in the homogeneous returns framework[8], are given by[9]:

$$Y_i = \delta + X_i\beta + D_{1i}\alpha_1 + D_{2i}\alpha_2 + D_{3i}\alpha_3 + \mu_i \qquad (1)$$

where X are the observed covariates[10], D_{1i} is a dummy taking the value of one if individual has attended Vocational high school and zero otherwise; D_{2i} is a dummy taking the value of one if the individual has attended Professional high school and zero otherwise; D_{3i} refers to General high school; $\alpha_{1i,}$ α_{i2} and α_{i3} are the average effects of the school types on earnings compared to the state of having attended Teaching/art high school in this case. The error term u_i is assumed to be independently and identically distributed across individuals $E[u_i] = 0$. Selection bias in equation (1) arises because of a stochastic relationship between D_{1i} , $D_{2i,}$ D_{3i} and $u_{i,}$ that is $E[u_i \mid D_{1i,} D_{2i,} D_{3i,} X_i] \neq 0$. Moreover, the treatment assignment is modelled by a latent variable D^*_{ij} as follows $D^*_{ij} = W_i\eta_j + \varepsilon_i$, where W_i are covariates assumed to affect the selection into a treatment, and j=1,2,3,4 indexes the various treatments. Enrolment at the high school types is thus defined by the following:

---

[8] In the homogeneous returns framework, the rate of return to a given high school type j is the same across individuals; that is, $\alpha_{ij=}\alpha_j$ for all individuals i.

[9] The log wage premia are referred to employed graduates. In this section, I don't address the question of what determines the probability of employment.

[10] It is implicitly assumed that the observables X are exogenous in the sense that their potential values do not depend on treatment status or equivalently that their potential values for the different treatment states coincide. Natural candidates for X that are not determined ar affected by treatments D are time-constant factors, as well as pre-treatment characteristics.

$D_{1i} = D_{2i} = D_{3i}=0$	iff	$D_{0i}^* = \max(D_{0i}^*, D_{1i}^*, D_{2i}^*, D_{3i}^*)$,
$D_{1i} = 1, D_{2i} = D_{3i}=0$	iff	$D_{1i}^* = \max(D_{0i}^*, D_{1i}^*, D_{2i}^*, D_{3i}^*)$,
$D_{1i} =D_{3i} = 0, D_{2i}=1$	iff	$D_{2i}^* = \max(D_{0i}^*, D_{1i}^*, D_{2i}^*, D_{3i}^*)$,
$D_{1i} = D_{2i} = 0, D_{3i}=1$	iff	$D_{3i}^* = \max(D_{0i}^*, D_{1i}^*, D_{2i}^*, D_{3i}^*)$,

Selection bias as described above can arise from two sources. First, the dependence between W_i and u_i is when the selection is on observables. If selection is assumed to be on observables, and furthermore, the dependence between W_i and u_i is assumed to be linear, the vector W_i may be included into the equation of outcome (1) to obtain estimates of the average treatment effects[11]. Thus to obtain unbiased[12] estimates of the treatment effects the following equation is estimated:

$$Y_i = \delta + X_i\beta + W_i\gamma + D_{1i}\alpha_1 + D_{2i}\alpha_2 + D_{3i}\alpha_3 + \upsilon_i \quad (2)$$

Second, there may be a dependence between ε_i and μ_i, usually referred to as selection on unobservables. Performing OLS estimation does not control for different sources of bias due to unobservables[13] (Blundell, Dearden and Sianesi, 2003): (i) Ability bias. This derives due to the likely correlation between the δ intercept term (absolute advantage) and D_{ij}. If higher-ability or inherently more productive individuals tend to acquire more frequently general than technical education, the two terms will be positively correlated, inducing an upward bias in the estimated average return α_3; (ii) Measurement error bias. One can think of μ_i as including measurement error in the schooling variable D_{ij}. Note that since the high school variable D_{ij} is a dummy variable, measurement error will be non-classical. Kane, Rouse and Staiger (1999) show that OLS estimates may be biased and that it is not possible to place any a-priori general restrictions on the direction or magnitude of the bias of the estimator.

[11] This is called the linear control function approach, see e.g. Heckman and Robb (1985).
[12] However, if the true model contains higher-order terms of the X_s, or interactions between the various X_s, the OLS estimate of α_i would in general be biased due to omitted variables (mis-specification of the no-treatment outcome $X_i\beta$). Moreover simple OLS constrains the returns to be homogeneous, if by constrast the effect of high school type varies according to some of the X_s, the OLS estimate of α_i will not in general recover the ATT. These mis-specification issues are linked to the bias due either to non-overlapping support of the observables X or to mis-weighting the observations to control fully for the difference in the distribution of X over the common region.
[13] In the heterogeneous returns model, there is another source of bias: Returns bias. This occurs when the individual returns component is itself correlated with the schooling decision. The direction of this bias is not clear.

5.2 Matching estimator (PMS-ATT model)

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, the matching method tries to mimic ex post an experiment by choosing a comparison group from among 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 choosing the appropriate matching variables. In some ways, this mirrors the issue of choosing an appropriate excluded instrument in the control function approach. However, instruments do not make appropriate matching variables and viceversa [14] (Blundell, Dearden and Sianesi, 2003) .

Hence an alternative method to estimate school type premia is to compare occupational wages for individuals who graduated in one school type and decided to participate to the labour market with matched individuals who attended a different school and decide to participate to the labour market too. Following the language of among other Lechner (1999b), this multiple evaluation problem can be presented as follows.

Consider participation in (M+1) mutually exclusive treatments, denoted by an assignment indicator $D \in (0,\ldots\ldots M)$. In my case, in particular, I assume that an individual can choose among five different alternatives $D \in (0,\ldots.,4)$, which are: (0) any school type and non participation to the labour market, (1) Vocational high school and participation, (2) Technical high school and participation, (3) general high school and participation, and (4) Teaching/arts high school and participation. Hence even if decision making is sequential, I will consider a joint decision model with partial observability, where the decision of individuals to enrol at one of the available high school type and the one to participate to the labour market are not taken separately.

Denote, then, variables unaffected by treatments, often called attributes or covariates, by X. The outcomes of the treatments are denoted by $(Y^0,\ldots..,Y^M)$, and for any participant, only one of the components can be observed in the data. The remaining M outcomes are called counterfactuals. In my case the outcomes are occupational wages earned by high school graduates three years after graduation.

[14] Instruments should satisfy an exclusion condition in the outcome equation conditional on the treatment, whereas matching variables should affect both the outcome and treatment equations.

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 evaluation is to define the effect of treatment m compared to treatment 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:

$$\theta_0^{ml} = E(Y^m - Y^l \mid D = m) = E(Y^m \mid D = m) - E(Y^l \mid D = m) \quad (3)$$

θ_0^{ml} in equation (3) denotes the expected average treatment effect of treatment m relative to treatment l for participants in treatment m (sample size N^m). In the binary case, where m=1 and l=0, this is usually called "treatment-on-the-treated" effect. 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 (3) can be identified under the following assumptions:

- Stable-unit-treatment-value assumption (SUTVA): potential outcomes for an individual are independent of the treatment status of other individuals in the population (cross-effects and general equilibrium effects are excluded).
- Conditional independence assumption (CIA): 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 $(Y^0,……, Y^M) \perp D \mid X=x, \forall x \in X$. Put it in a different way, given all the relevant observable characteristics (X), when choosing among the available treatments, the individual does not base this decision on the actual outcomes of the various treatments.

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

$$0<P^m(X)<1, \text{ where } P^m(x)=E[P(D=m \mid X=x)], \ \forall\, m=0,1,\dots.M \quad (4)$$

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 I seek to make a comparison. Depending on the sample in use, this can be quite a strong requirement. If there are regions where the support of X does not overlap for the treated and non-treated groups, matching has in fact to be performed over the common support region; the estimated treatment effect has then to be redefined as the mean treatment effect for those treated falling within the common support.

In the binary case of two treatments, Rosenbaum and Rubin (1983) show that if CIA is valid for X, it is also valid for a function of X called the balancing score b(X), such that $X \perp D \mid b(X)$. The balancing score property holds even for multiple case:

$$(Y^0,\dots\dots\dots,Y^M) \perp D \mid X=x, \ \forall\, x \in X \ \text{ if } (Y^0,\dots\dots\dots, Y^M) \perp D \mid b(X)=b(x), \ \forall\, x \in X \quad (5)$$

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 binary case of two treatments, the balancing score with the lowest dimension is the propensity score $P^1(x) = E[P(D=1| X=x)]$. In the case of multiple treatments, a potential and quite intuitive balancing score is the M-dimensional vector of propensity scores $[P^0(x), P^1(x)\dots\dots\dots.P^M(x)]$.

To identify and estimate θ_0^{ml}, first of all I identify and estimate E(Y^m | D=m) by the sample mean. The conditional independence assumption implies that the latter part of equation (3), E(Y^l | D=m), is identified in large enough samples as:

$$E[\, E(Y^l \mid b(X), D=m) \mid D=m] = E(Y^l \mid D=m) \qquad (6)$$

To estimate (6), Imbens (1999) and Lechner (1999b)[15] 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,

$$E(Y^l \mid D=m) = E[\, E(Y^l \mid P^m(X), P^l(X), D=l) \mid D=m] \qquad (7)$$

As suggested by the theoretical framework, I assume that an individual can choose among five different alternatives m (0,….,4), as specified above, each of them providing a utility of $U_{im} = \eta_m^{'} X_i + u_{im}$ where u_{im}= ε_m+v_m is the unobserved components in these utilities (see equation 3.5). I decide to model this choice using a multinomial logit model[16], though it assumes the independence of irrelevant alternatives, an assumption which contrasts with what we have seen in the theoretical framework[17]. To check whether this assumption is too restrictive, I have applied binomial logit models to estimate the propensities for all five comparisons and I have also performed the Hausman and Mcfadden test.

Hence, the discrete choice model to estimate the propensities is:

[15] A detailed description of the matching algorithm is presented in the Appendix.
[16] The multinomial probit model would be a better specification in this case. However my dataset does not contain alternative-specific variables, i.e. variables affecting the utility of a specific alternative only, in order to formally identify the econometric model. As stated by Keane (1991) the multinomial probit without these exclusion restrictions may suffer from tenuous identification especially when considering the choice among a number of alternatives higher than three.

.

$$\Pr(D_i = m) = \frac{\exp(X_i \eta_m)}{\sum_{l=0}^{4} \exp(X_i \eta_l)} \qquad (8)$$

where i indexes the individual, X is the vector of attributes including family background characteristics (parents education) and academic performance at junior high school (according to the value added specification described in section 3) and m indicates the choice. As I have just said before, I consider a joint decision model with two decision functions observed jointly: the decision of individuals to participate to the labour market and their decision to enrol at one of the available high school type. In this way, I manage to take into account student self-selection both into school type and participation to the labour market.

6.1 Empirical Results

6.1.1 Estimation of the Propensity

The discrete choice model to estimate propensities is a multinomial logit model with five choice alternatives as suggested by the theoretical and methodological framework: participation to the labour market and either General (Licei) or Professional (Istituto Tecnico) or Vocational (Istituto Professionale) or Teaching/art (Magistrali, Istituti d'arte, altro) high school, and non participation to the labour market. The Professional high school and participation to the labour market is considered as the reference category in this case. The matching variables (X) included as covariates in the discrete choice model are the following: parents' education, father's occupation, gender, region of residence, year of birth, junior high school score.

The assumption of independence of irrelevant alternatives (IIA) underlying the multinomial logit model may be argued to be too restrictive in this context. To check whether this is the case I have applied binomial logit models to estimate the propensities for all five comparisons. The results are presented in Table 8. The estimated coefficients of the binomial and multinomial models are similar to each other, and thus the IIA assumption is considered to be sufficiently valid. I have also performed the Hausman and Mcfadden test excluding the Professional high school choice: this seems the most interesting restriction to impose because it is more likely to be correlated with the Vocational high school choice. Also this test does not reject the IIA assumption.

The results presented in Table 7 show that the significance of various explanatory variables differs across the three school types. For example, parents' education[18] and not generally father occupation seems to affect positively the decision to enrol at General high school, especially when both parents have university degree. In the Vocational high school case instead, both parents' education and father's occupation exhibit a high explanatory power: having a father employed in qualified jobs (retailer, entrepreneur, professional, teacher, white collar) or at least one parent with a university degree, influences negatively the decision to enrol at Vocational high school. Consistently with the results obtained in the main literature on the impact of family background on secondary schooling in Italy (Checchi et all 2006; Cappellari 2005), I found that variables related to family background have a substantial and significant impact on the high school choice, even after controlling for ability.

The dummy-variables for the junior high school score seem in general to have high significance, indicating that also individual ability, and not only parental background, select pupils into tracks. In particular, the higher is the grade the higher is the likelihood that pupils enrol at General high school than at Professional institute. The other way around if we consider Vocational and Teaching/art high school choices. Finally, there is evidence that female students are more likely to enrol at General and Teaching/art high school than Professional one.

Distribution of the predicted propensities are presented in Figure 2. In broad outline, a good model produces large differences of the mean predicted propensities across the various groups. This is the case for all the estimated propensities.

A correct estimation of the average treatment effects requires common support for the treatment and the comparison group, i.e. $0<P^m(x)<1$ for all m=0,….M. In practice, this implies that some observations are excluded from the sample, if the propensity distributions do not cover the exact same interval. In other words, an observation in the subsample m with an estimated propensity vector equal to $[p^0(x), p^1(x)\ldots\ldots\ldots p^M(x)]$ is excluded from the sample if any of these propensities is outside the distribution of the specific propensity in any other subsamples l. Due to this common support requirement, approximately 5000 observations are deleted leaving a sample size of nearly 13000.

[18] The biasedness of these coefficients is not so severe if we assume that junior high school grades a good proxy for contemporaneous unobserved family inputs and students' ability.

6.1.2 Matching

In the binary case of one treatment, the sub-sample of non-treated consists generally of a large number of observations, and thus each comparison unit can be used only once. In the multiple case this is not meaningful, since pair-wise comparisons are done across all sub-samples, and for some comparisons, the potential comparison group is much smaller than the treatment group. Thus matching is done with replacement, i.e. each comparison unit may be use more than once given that it is the nearest match for several treated units. The covariance matrix for the estimates of average effects, suggested by and presented in Lechner (1999b), pays regard to the risk of "over-using" some of the comparison units: the more times each comparison is used, the larger the standard error of the estimated average effect.

A detailed description of the matching algorithm is presented in the Appendix 2. The pair-wise matching procedure is carried through altogether 12 times. 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.

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 show that, in general, the match quality is almost satisfactory for the reported model specifications, and thus I consider the condition $X \perp D \mid b(X)$ to be sufficiently fulfilled.

6.1.3 Average treatment effect on the treated

The last column of Table 10 reports the results for the twelve various treatment on the treated effects. I use Mahalanobis matching with replacement so that in some cases a non-treated individual provides the closest match for a number of treated individuals, whereupon they feature in the comparison group more than once.

I seek to ensure the quality of my matches by setting a tolerance when comparing propensity scores. I do so by imposing three different calipers (0.01, 0.001 and 0.0001): where propensity score of a treated individual falls beyond these bounds for a near comparator, the treated individual remains unmatched. This second means of enforcing common support results in the discarding of further members from our analysis.

First, let us compare General high school to the Professional and Vocational high school types: the effects on monthly wages three years after graduation are negative and statistically significant, irrespective of the specification of the caliper employed. Hence, according to PMS-ATT estimates, General high school graduates have a statistically significant negative wage premium with respect to Professional and Vocational high school graduates. This is either because of skill content of General high school curriculum, less valued by employers compared to Professional/Vocational ones or because of bad signalling, i.e. the labour market does not expect General school students to be engaged in job search activities and interpret their participation as a sign of low ability. On the other hand, there seem to be a positive wage differential between General and Other high school graduates. However in this case the ATT is no statistically significant in all caliper specifications.

Second, comparisons of Professional and Vocational institutes versus General high schools confirm that the formers perform better than the latter in terms of wages three years after graduation. The coefficients are always statistically significant in these two cases.

Finally there is no evidence from the PMS-ATT estimates that Vocational, Professional and Teaching/art high schools exhibit wage premium with respect to each other.

6.1.4 Is it plausible to assume Conditional Independence?

In the literature of economics of education a lot of studies on school choice pointed out that parents' education, father's occupation, student's ability, gender and region of residence are important factors in determining which high school type an individual will enrol at. These factors are also likely to influence the future labour market outcome, and thus, in order for conditional independence to be plausible, they should be included in the estimation of the propensities.

The importance of parental background for the children's educational choices and attainments is emphasized in various studies, starting with Haveman and Wolfe (1995). Examples of more recent studies that all point to parents' education as one of the most essential factors to be controlled for in measuring the effect of education on earnings are Figlio (2000), Blundell et al (2003), Dustmann (2004), Cappellari (2004), Checchi et al (2006) and Hanushek et al (2005).

Both parents education, father's occupation when the child was 14^{th} years old and academic performance prior to secondary school (junior high school score) are all included in the data available for this study. Moreover, the data set provide detailed information on the other personal characteristics, such as age, gender and region of residence. Information is missing on some specific indicators of students' cognitive ability and other important family characteristics (such as household income). However, as we have already seen in section 3, we can consider on one hand the junior high school score as a good proxy for unobserved student's abilty and on the other parents' education and occupation as good indicator of family social class and wealth.
Hence the available data include much, but not necessarily all information on factors which affect the selection and the outcome. The crucial question –that is left to the reader to decide- is whether there is sufficient information to justify the conditional independence assumption. Later on, in section 6.3, I will apply different methods to the same problem and comparing the results. In short, I find that different methods produce somewhat different estimates for the high school type effects, but the sign of the effects is essentially the same across methods.

6.2 OLS model.

Results for the OLS regression are presented in table 6. The set covariates includes the variables used to estimate the propensity scores. Comparison of the results in Table 10 shows that OLS on the one hand and matching on the other produce different estimates. In this specific case, the matching estimates for the ATTs are generally lower than OLS ones.
Both matching and OLS deal with observables only; matching however, also offers simple and effective ways of assessing ex post the quality of a matched comparison group in terms of the observables of interest. Semi-parametric methods such as matching thus force the researcher to compare only comparable individuals. In a given application one would expect little bias for ATT from simple OLS vis-à-vis matching if there is:

1. no common support problem;
2. little heterogeneity in treatment effects according to X or, alternatively, all the propensity scores are small (in particular, less than 0.5, which would make the

weighting scheme of OLS proportional to the one for the matching estimator ATT- see Angrist (1998));

3. no serious mis-specification in the no-treatment outcome.

The common support restriction is generally binding for the ATT (about 40% of the observations falls outside of the common support). Moreover the shares of the General, Professional, Vocational and Teaching/art propensity scores larger than 0.5 are generally high. Hence even if the specification of no-treatment outcome is reasonably correct, I would expect matching and simple OLS to produce different estimates of the ATT. Note that matching dominates simple OLS a priori. Matching can quickly reveal the extent to which the treated and non-treated groups overlap in terms of pre-treatment variables, it offers easy diagnostic tools to assess the achieved balancing and it relieves the researcher from the choice of the specification of the no-treatment outcome.

6.3 Sensitivity analysis

This section report the robustness of the results presented in the previous section, referred to as the "main analysis" or "the main results".

Firstly, the sensitivity of the results to the methodology used to estimate the average treatment effect of school type on earnings. 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, the costs of using matching are not so unimportant. On one hand the assumption of conditional independence is not only very strong but also impossible to test. On the other, even though one does not have to specify the outcome model, there are several other decisions to make concerning among others, the specification of the discrete choice model, the criterion of matching, and the definition of common support. Hence, in this section I introduce another approach for determining the average treatment effect on the population and relate it to the method of propensity score matching and the results presented in the previous section. In particular, I apply the polychotomous selectivity model introduced by Lee (1983) to investigate the existence of unobserved heterogeneity.

Secondly, I investigate whether there is some heterogeneity in the treatment effects between women and men. Thirdly, I will make a robustness check on the assumption of partial observability used to estimate the propensities, too see if participation and employment are post-treatment effects rather an individual decision, once high school type is chosen. Finally, I extend the analysis to the 2004 edition of the survey on job opportunities for the 2001 Italian high school graduates to see if the results are robust to the sample used.

6.3.1 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 with 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[19]. Like all these selectivity models, the Lee model is designed to adjust for both observed and unobserved selection bias. Thus, it does not require the conditional independence assumption to be valid. However, it rests on other strong assumptions, among them linearity in the outcome variable and joint normality in the error terms. Consider the following model:

$$y_1 = x\beta_1 + u_1$$
$$y_m^* = z\gamma_m + \eta_m, \quad m=0,\ldots..,M \qquad (6.1)$$

where the disturbance μ_1 is not parametrically specified and verifies $E(u_1 \mid x, z) = 0$ and $V(u_1 \mid x,z) = \sigma^2$; m is a categorical variable that describes the choice of an economic agent among M[20] alternatives based on "utilities" y_m^*. 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 and participation to the labour market by way of some excluded instruments. The choice of an appropriate instrument z, like the choice of the appropriate conditioning set x for matching or OLS, boils down to an untestable prior judgement. My data set does

[19] The main shortcoming of the Lee approach compared to the one presented by Dubin and Mcfadden is that it contains relatively restrictive assumptions on the covariance between the error term ε and μ (see below for more details).
[20] As in the estimation of the propensities in the matching framework, I choose a joint decision model with partial observability. Hence the choice alternatives are: any high school and non participation to the labour market (m=0), participation and enrolment at vocational high school (m=1), participation and enrolment at professional high school (m=2), participation and general high school (m=3), participation and Teaching/art high school (m=4).

not contain the potential excluded variables used in the related literature that may determine assignment to school types but, conditional on the x_s, be excluded from the earnings equation, such as parents' interest in child education and adverse financial shocks hitting the child's education at age 13. Hence I decided to use as "instruments" the number of siblings, the grandfather's education because they seem to be significant determinants of the secondary school choice (conditional on the full set of controls X), with individual F-values ranging from 20.47 to 26.47, and not of the early earnings.

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

$$y_1^* > \max_{\neq 1} (y_j^*) \qquad (6.2)$$

Define $\varepsilon_1 = \max_{m\neq 1}(y_m^* - y_1^*) = \max_{m\neq 1}(z\gamma_m^* + \eta_m - z\gamma_1 - \eta_1)$ (6.3). Under definition (6.3), condition (6.2) 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 with:

$$P(\varepsilon_1 < 0 \mid z) = \frac{\exp(z\gamma_1)}{\sum_m \exp(z\gamma_m)}$$

Based on this expression, 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 (6.1). Because of this, least squares estimates of β_1 would not be consistent.

Define Γ as follows:

$$\Gamma = (z\gamma_1, \ldots\ldots\ldots, z\gamma_M)$$

Generalizing the Heckman (1979) model, bias correction can be based on the conditional mean of u_1:

$$E(u_1 \mid \varepsilon_1 < 0, \Gamma) = \iint_{-\infty} \frac{u_1 f(u_1, \varepsilon_1 \mid \Gamma)}{P(\varepsilon_1 < 0 \mid \Gamma)} d\varepsilon_1 du_1 = \lambda(\Gamma)$$

where $f(u_1, \varepsilon_1 | \Gamma)$ is the conditional joint density of u_1 and ε_1. For notational simplicity, call P_k the probability that any alternative k is preferred:

$$P_k = \frac{\exp(z\gamma_k)}{\sum_m \exp(z\gamma_m)}$$

Given that the relation between the M components of Γ and the M corresponding probabilities is invertible, there is a unique function μ that can be substituted for λ such that:

$$E(u_1 \mid \varepsilon_1 < 0, \Gamma) = \mu(P_1, ..., P_M)$$

Therefore, consistent estimation of β_1 can be based on either regression:

$$y_1 = x_1\beta_1 + \mu(P_1, ..., P_M) + \varpi_1 = x_1\beta_1 + \lambda(\Gamma) + \varpi_1 \quad (6.4)$$

where ϖ_1 is a residual that is mean-independent of the regressors. As argued by Dahl (2002), semi-parametric estimation of this model would have to face the "curse of dimensionality". Whenever the number of alternatives is large, it implies the estimation of a very large number of parameters, which rapidly makes it intractable for practical implementation. As result, restrictions over λ(Γ), are required. 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 however achieved at the cost of fairly restrictive assumptions.

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 $J_{\varepsilon 1}(.|\Gamma)$ are linearly related with correlation ρ_1 (this holds in particular if they are bivariate normal). That is:

- $corr(u_1, J_{\varepsilon 1}(\varepsilon_1 | \Gamma)$ does not depend on Γ (Lee's correlation assumption).
- $E(u_1 | \varepsilon_1, \Gamma) = -\sigma\rho_1 J_{\varepsilon 1}(\varepsilon_1 | \Gamma)$ (Lee's linearity assumption).

Then, the expected value of the disturbance term u_1, conditional on category 1 being chosen, is given by:

$$E(u_1 | \varepsilon_1 < 0, \Gamma) = -\sigma\rho_1 \frac{\phi(J_{\varepsilon 1}(0|\Gamma)}{F_{\varepsilon 1}(0|\Gamma)}$$

with Φ the standard normal density. Under this hypothesized form for λ(Γ), a consistent estimator of β_1 is obtained by running least squares on the following equation:

$$y_1 = x_1\beta_1 - \sigma\rho_1 \frac{\phi(J_{\varepsilon 1}(0|\Gamma)}{F_{\varepsilon 1}(0|\Gamma)} + w_1 \quad (6.5)$$

Two-step estimation of (6.5) is thus implemented by first estimating the (γ_j)'s in order to form $\frac{\phi(J_{\varepsilon 1}(0|\Gamma))}{F_{\varepsilon 1}(0|\Gamma)}$ and then by including that variable in equation (6.5) to estimate consistently β_1 and $\sigma\rho_1$ by least squares.

Results from the empirical application are presented in tables 11 and 11.1 . The sample is the same as in Section 6.2. The set of X variables in the wages equation is limited to: parents' education, father's occupation, sex, region of residence, year of birth, high school score and junior high school score. Number of siblings and the dummy related to grandfather's education are thus assumed to affect selection into high school type but not wages. The results in Table 11.1 show that including the selection adjustment terms in the equation for earnings produces higher estimates in absolute value of the ATTs compared to the matching ones. But as in the matching framework, Vocational and Technical high school graduates seem to have a positive wage premium with respect to graduates from General high school. However it is not straightforward to draw conclusions about the existence of unobserved heterogeneity, because the parameter estimates for selection adjustment terms are not statistically significant at the 5% level. Hence, there appears to be some evidence[21] suggesting that there are enough variables to control directly for selection on unobservables. In other words, I could have found some evidence that OLS and matching with available set of Xs are not subject to selection bias.

6.3.2 Separate analysis of women and men

In order to examine whether there is some heterogeneity in the treatment effects between women and men, the sample is divided by sex, and the matching procedure[22] from section 6.1 is applied to analyse the average treatment effects conditional on sex. The results are presented in table 12. In brief, there is considerable heterogeneity between the sexes, which is discovered by comparing the relative treatment effects expressed in percentage. The precision of the estimates is in general better for women than for men, which might be explained by the larger sample size for men. The effects of General high school on earnings are more negative for women than for men. Consequently general education appears as significantly worse in terms of early labour market outcomes for women. Concerning vocational/professional education, the opposite holds: Vocational high school for example seems to be more favourable for women than for men. Hence vocational/professional education seems to have been a better choice for women aiming at employment after graduation. Generally women have fewer job opportunities than men do, but this type of education seem to help them ease their way into the labour market.

6.3.3 Partial observability versus sample separation information.

In this paragraph, I check whether the decision model with partial observability, used to estimate the propensities in section 6.1, is too restrictive in this context and does mask the impacts of high school type on other aspects of transition to the labour market, namely participation and employment. To see if this is the case, I firstly estimate[23], on the entire sample of graduates, the relative probability of participation to the labour market by high school type. Taking into account only those individuals who decide to participate, I

---

[21] 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.

[22] I choose to specify a 0.0001 caliper matching in order to compare only the most comparable individuals, even if imposing a very restrictive matching increases the standard errors of the estimated treatment effects.

[23] see 22.

estimate the relative probabilities of employment. And at last, after dropping out from the sample those individuals who are unemployed, I re-estimate the average matching treatment on the treated effects on earnings correcting only for student self selection into high school type[24]. Table 13 report the distribution of the participation and employment rates by high school type. On one hand, we can see that most of Vocational and Technical high school graduates decide to participate to the labour market (about 90% and 75% respectively), of whom nearly 60% are employed. On the other it is interesting to see that only 32% of General high school graduates participate to the labour market after graduation and only 27% of them have found a job.

Table 13.1 reports the results for the twelve various treatment on the treated effects[25]. As before, each estimated effect is reported in relative terms. First, let us compare the high school types with respect to participation to the labour market. Both Vocational and Professional schools have a positive and statistically significant effect on participation probability: Vocational and Professional high school graduates have nearly a 25% higher probability to participate to the labour market with respect to General high school ones. The same probability is 14% higher for Teaching/art high school graduates. These results are consistent with the main purpose of a school system stratified into specialized tracks, like the Italian one: to allocate differently pupils either to labour market oriented or academic track in relation to their needs and tastes. After having shown that the vocational/technical tracks enhance labour market participation, let us verify if they also improve the chances of getting a job. Focusing on the second column of table 13, we can see that among high school graduates who decide not to go on further education, Technical and Professional high school ones are the most advantaged in terms of employment probability. Finally, the average treatment effects on earnings conditional on employment confirm the results obtained taking into account student self-selection both into participation to the labour market and into high school type. General high school graduates have a negative wage differential with respect to Professional and Vocational high school ones. There is not any earning premium in relation to Teaching/art high school with respect to the other high school types and there is no evidence that Vocational and Professional high schools exhibit wage differential with respect to each other. However, the treatment effects on earnings, conditional on participation and employment, are less

[24] Hence these effects are to be interpreted conditional on participation and on employment.

[25] The covariate balancing indicators before and after matching always indicate that the match quality is almost satisfactory. Details are available on request.

precisely estimated and are generally different in absolute value because they don't take into account of the endogeneity of participation decision.
Hence, the results from the sensitivity analysis suggest that Italian secondary school system differentiated into different school tracks is generally efficient: vocational and technical education increases not only participation to the labour market but also employment probability and early earnings.

6.3.4 Estimation of the ATTs using the sample of 2001 Italian high school graduates.

In this section, the analysis of the differences in early earnings by high school type is accomplished by using the most recent survey on job opportunities for the Italian high school graduates. The sample[26] consists of 20,408 records and it has been reduced to 18,548 by eliminating those who were employed and their job while at high school. The log net monthly earning is available for nearly 9300[27] from which I dropped the extreme observations.
To estimate the propensity scores I used exactly the same specification and matching variables as the ones of the main section. The distribution of the predicted propensities are presented in figure 4. The common support requirement implies the elimination of nearly 1,500 individuals, leaving a sample size of 17,129. Finally, the covariates are sufficiently balanced by the reported model specification , according to the mean absolute bias[28].
Table 14 reports the results for the twelve various treatment on the treated effects. Comparing these results to the previous ones shows that using the samples of 1995 graduates and of 2001 graduates produce fairly similar estimates of the average treatment on the treated effects both in sign and absolute value for the different caliper specifications. General high school graduates, once again, exhibit a negative wage differential with respect to both Technical and Vocational high school graduates.

[26] See in the appendix for main descriptive statistics.
[27] Its distribution is given in figure 2.
[28] Results about mean absolute bias before and after matching are available on request.

7. Concluding remarks.

This work is focused on the gains from specialization implied by a stratified educational system. To prove the efficiency of a system organized into tracks, I provide empirical evidence on differences in early occupational labour market outcomes between graduates in General (Licei), Vocational (Istituti Professionali), Professional (Istituti Tecnici) and Teaching/art school (Istituti d'arte/Istituto magistrale). This paper, in particular, estimates relative log-wage premia of employed graduates by high school type three years after graduation. The results obtained from a standard OLS approach are contrasted with estimates from propensity score matching technique which correct for selectivity through observable characteristics. There appears to be some evidence suggesting that there are enough variables to control directly for selection on unobservables, so I conclude that matching estimates with available set of Xs are not subject to further selection bias. I found that, irrespective of the estimation technique employed, Vocational and Technical high school graduates have generally a positive wage premium with respect to graduates from General and Teaching/art high school types. The reasons for these results could be: (i) the skill content of General and Teaching/art high school curriculum may be less valued by employers compared to Professional/Vocational schools; (ii) the labour market does not expect especially General high school students to be engaged in job search activities and interpret their participation as a sign of low ability. The same results are obtained using the sample of 2001 Italian high school graduates. Dividing the sample by sex and estimating the treatment effects conditional on sex show that the negative effect of academic/general education on earnings seem to be greater for women than for men. Moreover the estimation of the model where participation and employment are considered as post-treatment effects of the high school type suggest that vocational and technical education increases not only early earnings but also participation to the labour market and employment probability.

Hence it seems that, contrary to the policy stance of the World Bank that favours general secondary education, for those youths proceeding directly to the labour market after leaving high school technical/vocational education is superior to general one in terms of early earnings and employment probabilities. These findings could suggest that all the recent attempts of reforms, trying to empty technical education of all meaning, were not the best way to renew the secondary school system. The proposal of reform of previous government[29] was, in fact, to reduce the number of tracks to two, introducing the concept of dual channel. The new system, in particular, would have only provided the “Lyceums”, propedeutic to university (the former general schools and only few of the previous technical and vocational schools) and the new “vocational” track (the most of the previous technical and vocational schools) with less time devoted to actual instruction and consequently fewer technical skills acquired. Moreover the instruction provided at the lower vocational track would be at a lower level of complexity inhibiting further students’ chances of success in school and of continuing on to college.

Hence the introduction of this dual channel could have reduced the efficacy of the Italian tracked system and could have increased the risks of social segmentation.

[29] See in the appendix for more details.

References:

Aghion et al (1999), " Technical change, inequality and the labor market", NBER working paper no.7800, Cambridge, MA.

Arum and Shavit (1995), "Secondary vocational education and the transition from school to work"; Sociology of Education, vol.68, 3, pp. 187-204

Andrews M., Bradley S. and Stott D. (2002): The school-to-work transition, skill preferences and matching, Economic Journal, 112, C201-C219.

Bennell P. (1996): General Versus Vocational Secondary Education in Developing Countries: A Review of the Rates of Return Evidence; Journal of Development Studies, Vol.32(2):230-247.

Bertocchi, G. and Spagat M. (2003): The evolution of Modern Education System, in Journal of Development Economics.

Bratti M. (2004): Social class and undergraduate degree subject in the UK, IZA discussion paper No 1979.

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.

Bourguignon F., Fournier M., and Gurgand M. (2001): Selection Bias Correction based on the Multinomial Logit Model, CREST Working Paper, 2001.

Card, D. (1999): The causal effect of education on earnings, in O. Ashenfelter and D.Card, Handbook of Labor Economics, vol.3, Amsterdam: Elsevier-North Holland.

Cappellari L. (2004): High school types, academic performance and early labour market outcomes, IZA Working Paper.

Checchi D. (2000); "University Education in Italy", mimeo.

Checchi D. and Flabbi L. (2005): Intergenerational Mobility and Schooling decisions, mimeo.

Dustmann C. (2004): Parental background, secondary school track choice, and wages, Oxford University Press 56, 209-230.

Figlio, D. and Page, M. (2000), School Choice and The Distributional Effects of Ability Tracking: Does Separation Increase Equality? NBER Working Paper 8055.

Franz W, Pohlmeier W, and Zimmermann V, (2000): Young and Out in Germany: On youths chances of labor market entrance in Germany, in Youh Employment and Joblessness in Advanced Countries, edited by Blanchflower D; and Freeman R; University of Chicago Press, Chicago.

Heckman J., Ichimura, H., and Todd, P. (1998): Matching as an econometric evaluation estimator; Review of Economic Studies, 65, 261-94.

Haveman R. and Barbara Wolfe (1995): The determinants of children's attainments: a review of methods and findings; Journal of Economic Literature vol. XXXIII (December 1995), pp. 1829-1878.

Imbens, G. (1999): The role of Propensity Score in Estimating Dose-Response Functions, NBER working paper TO237.

Lechner, M. (1999a): Earnings and Employment Effects of Continuous Off-the Job Training in East Germany After Unification, Journal Of Business & Economic Statistics, 17, 74-90.

Lechner, M. (1999b): Identification and estimation of causal effects of multiple treatment under the conditional independence assumption, Discussion paper 9908, University of St. Gallen.

Lee, L. F. (1983): Generalized Econometric Models with Selectivity, Econometrica 51(2): 507-512.

Maddala (1983): Limited Dependent Variables, Cambridge University Press.

Margolis D N et al (1999): The role of early career experiences in determining later career success: An international comparison, unpublished paper, TEAM-Universite de Paris Pantheon-Sorbonne.

Margolis, D.N. and Simonnet, V. (2003): Educational Track, Networks and Labor Market Outcomes, Institute for the Study of Labour, Bonn, Discussion Paper 699.

Meghir, C. and Palme (2004): Educational Reform, Ability and Parental Background, The American Review.

Mincer, J. (1974): Schooling, Experience and Earnings, National Bureau of Economic Research. New York, NY: Columbia University Press.

Neumann, G. R. (1999) "Search Models and Duration Data" in: Pesaran, M.H. and Schmidt, P., Handbook of Applied Econometrics, vol. II, Oxford: Blackwell.

Newman, S. and A. Zinderman (1991): Vocational Schooling, Occupational Matching and Labor Market Earnings in Israel. Journal of Human Resources, Vol.26(2): 256-281.

Psacharopoulos, G. (1987): To Vocationalize or not to Vocationalize: That is the Curriculum Question, International Review of Education, Vol. 33(2): 187-211.

Psacharopoulos, G. (1994): Returns to Investment in Education: A Global Update. World Development, Vol. 22(9): 1325-1343.

Poirier (1980): Partial observability in Bivariate Probit Models. Journal of Econometrics 12: 209-217

Rubin D.B. (1986): Discussion of Statistics and causal inference by Holland, Journal of The American Statistical Association, 81, 961-962.

Sianesi, B. et al (2001): Evaluating the impact of education on earnings in the UK: models, methods and results from the NCDS, Institute for Fiscal Studies WP03/20.

Sianesi, B. (2002): Essays on the evaluation of social programmes and educational qualifications, PhD thesis, University College London.

Sianesi B. (2004): An evaluation of the Swedish system of active labour programmes in the 1990s, Review of Economics and Statistics, 86, 133-155.

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

Tansel A. (1994): Wage Employment, Earnings and Returns to Schooling for Men and Women in Turkey, Economics of Education Review, Vol. 13(4): 305-320.

Trost R. and L.F. Lee (1984): Professional Training and earnings: A Polycotomous Model with Selectivity, Review of Economics and Statistics, Vol.66: 151-156.

Woessman L., Hanushek, E. (2005): Does Educational Tracking Affect Performance and Inequality? Difference-in-Difference Evidence across Countries", Economic Journal.

Table 1: Type of institute attended by high school graduates of 1995.

High school types	Gender		
	male	female	Total
Vocational	45.81	54.19	14,30
Professional	57.11	42.89	44,57
General	42.97	57.03	10,28
Teching/art school	11.65	88.35	30,86
Total	46.48	53.52	100

Table 2: Distribution of region of residence by high school type (%).

High school types	Region of residence			
	nord ovest	nord est	centro	sud
Vocational	21.35	20.22	23.3	35.13
Professional	23.62	17.16	22.07	37.16
Teaching/art	22.85	15.36	25.62	36.18
General	15.06	13.64	23.01	48.29
Total	22.18	16.68	23.43	37.71

Table 3: Distribution of high school grades by high school type (%).

High school types	High school grades			
	36-41	42-47	48-53	54-60
Vocational	37.61	33.82	17.37	11.21
Professional	38.64	30.8	17.42	13.13
Teaching/art	35.71	31.84	19.65	12.79
General	26.59	28.16	23.77	21.47
Total	38.84	32.39	18.04	10.73

Table 4: Distribution of junior high school grades by high school type (%).

High school types	Junior high school grades			
	Sufficient	Good	Very Good	Excellent
Vocational	65.91	25.21	7.07	1.81
Professional	34.96	33.21	20.28	11.55
Teaching/art	40	32.07	17.04	10.89
General	10.73	20.38	27.47	41.42
Total	32.17	27.9	20.37	19.57

Table 5: Distribution of Parental education by high school types.

High school types	Parental Education						
	level 1	level 2	level 3	level 4	level 5	level 6	level 7
Vocational	25.93	20.21	25.79	19.61	6.83	1.32	0.31
Professional	18.04	15.51	25.03	21.82	14.28	3.98	1.33
Teaching/art	20.06	15.25	22.28	20.52	14.09	5.78	2.02
General	6.23	7.92	11.95	19.75	25.34	17	11.8
Total	15.67	13.78	20.78	20.73	16.67	7.86	4.52

Note: level 1: both parents elementary school; level 2: at least one parent junior high school; level 3: both parents junior high school; level 4: at least one parent high school; level 5: both parents high school; level 6: at least one parent university; level 6: both parents university.

Table 6: OLS (uncorrected) Log Monthly Earnings Equation Estimates.

	Professional as ref		General as ref		Vocational as ref		Teaching/art as ref	
Variables	**Coef**	**P-value**	**Coef**	**P-value**	**Coef**	**P-value**	**Coef**	**P-value**
High school types[1]								
	-0.18683	0	0.18683	0	-0.168577	0	-0.0485778	0.231
	-0.018253	0.328	0.168577	0	0.018253	0.328	0.1199992	0
	-0.1382522	0	0.0485778	0.231	-0.1199992	0	0.1382522	0
Parents' education:								
At least one parent: junior high school	0.0072637	0.752	0.0072637	0.752	0.0072637	0.752	0.0072637	0.752
Both parents:junior high school	0.0385649	0.092	0.0385649	0.092	0.0385649	0.092	0.0385649	0.092
At least one parent: high school	-0.0142447	0.541	-0.0142447	0.541	-0.0142447	0.541	-0.0142447	0.541
Both parents: high school	0.0195366	0.516	0.0195366	0.516	0.0195366	0.516	0.0195366	0.516
At least one parent: university	-0.0852602	0.157	-0.0852602	0.157	-0.0852602	0.157	-0.0852602	0.157
Both parents: university	0.0515848	0.585	0.0515848	0.585	0.0515848	0.585	0.0515848	0.585
Father's occupation:								
Retailer	0.0734032	0.017	0.0734032	0.017	0.0734032	0.017	0.0734032	0.017
Craft	0.0244691	0.329	0.0244691	0.329	0.0244691	0.329	0.0244691	0.329
Farmer	0.0794764	0.063	0.0794764	0.063	0.0794764	0.063	0.0794764	0.063
Entrepreneur	0.0985584	0.009	0.0985584	0.009	0.0985584	0.009	0.0985584	0.009
Professional	0.0360124	0.439	0.0360124	0.439	0.0360124	0.439	0.0360124	0.439
Manager	-0.019336	0.755	-0.019336	0.755	-0.019336	0.755	-0.019336	0.755
Teacher	0.0247897	0.759	0.0247897	0.759	0.0247897	0.759	0.0247897	0.759
White collar high level	0.056579	0.022	0.056579	0.022	0.056579	0.022	0.056579	0.022
White collar low level	0.1018484	0.045	0.1018484	0.045	0.1018484	0.045	0.1018484	0.045
Blue collar high level	0.0583145	0.002	0.0583145	0.002	0.0583145	0.002	0.0583145	0.002
Female	-0.2203913	0	-0.2203913	0	-0.2203913	0	-0.2203913	0
Region of residence:								
Centre	-0.1043382	0	-0.1043382	0	-0.1043382	0	-0.1043382	0
North-east	0.0587447	0.002	0.0587447	0.002	0.0587447	0.002	0.0587447	0.002
South	-0.2682738	0	-0.2682738	0	-0.2682738	0	-0.2682738	0
Year of birth:								
born in 1976	-0.0160632	0.339	-0.0160632	0.339	-0.0160632	0.339	-0.0160632	0.339
born after 1976	-0.0870912	0.022	-0.0870912	0.022	-0.0870912	0.022	-0.0870912	0.022
High school score	0.0030733	0.061	0.0030733	0.061	0.0030733	0.061	0.0030733	0.061
Junior high school score								
good	0.0123857	0.516	0.0123857	0.516	0.0123857	0.516	0.0123857	0.516
very good	0.0147877	0.541	0.0147877	0.541	0.0147877	0.541	0.0147877	0.541
excellent	-0.013066	0.692	-0.013066	0.692	-0.013066	0.692	-0.013066	0.692
const	7.183488	0	6.996658	0	7.165235	0	7.045236	0

(1) see table 10 for more details about OLS ATEs

Table 7 Results from the multinomial logit estimates: estimates of the propensity.

Variables	Non-participation		General		Vocational		Teaching/art	
	Coef	P-value	Coef	P-value	Coef	P-value	Coef	P-value
Parents' education:								
At least one parent: junior high school	0.2209145	0.054	0.1710171	0.345	0.1142147	0.308	-0.1587096	0.272
Both parents:junior high school	0.166492	0.109	0.1341685	0.399	-0.2023947	0.041	-0.3049305	0.021
At least one parent: high school	0.5758796	0	0.7803745	0	-0.1952183	0.099	-0.1111994	0.47
Both parents: high school	1.014286	0	1.186371	0	-0.5655062	0	-0.1509933	0.429
At least one parent: university	1.794231	0	1.873517	0	-0.454227	0.084	0.5085662	0.125
Both parents: university	2.651465	0	2.458901	0	-0.1118803	0.895	0.5829115	0.241
Father's occupation:								
Retailer	0.2270345	0.069	-0.0011783	0.995	-0.6379906	0	-0.0913656	0.617
Craft	0.1497498	0.285	0.0705111	0.715	-0.2917418	0.031	-0.1589642	0.374
Farmer	0.0763965	0.72	0.232551	0.397	-0.1091033	0.566	-0.1753699	0.527
Entrepreneur	0.4580652	0.004	0.4861746	0.039	-0.7528321	0.002	-0.4592528	0.148
Professional	0.6204626	0	0.1842836	0.417	-0.6418241	0.024	0.0495102	0.893
Manager	0.9863064	0	1.037885	0	-0.2944352	0.363	0.0559158	0.884
Teacher	0.5963791	0.008	0.7008022	0.009	-0.8242285	0.072	0.1955839	0.584
White collar high level	0.1016695	0.346	0.284484	0.048	-0.4242615	0.003	-0.1569907	0.395
White collar low level	0.1220448	0.312	0.089164	0.597	-0.3279672	0.039	0.0173073	0.929
Blue collar high level	-0.2070314	0.032	-0.1989739	0.182	0.0701284	0.483	-0.0774971	0.559
Female	0.3181976	0	0.6087632	0	0.5700264	0	2.311289	0
Region of residence:								
Centre	0.0858352	0.315	0.2249036	0.053	0.0183889	0.872	0.5867081	0
North-east	-0.0455787	0.625	-0.3175752	0.02	0.3453321	0.002	0.4380234	0.005
South	-0.063566	0.432	0.4070615	0	-0.0320077	0.761	0.5503575	0
Year of birth:								
born in 1976	0.6650461	0	0.3792894	0	0.2617688	0.001	0.0233406	0.856
born after 1976	1.316593	0	0.9346846	0	-0.6084018	0.014	2.248443	0
Junior high school score								
good	0.1396181	0.091	0.2247437	0.083	-1.075731	0	-0.6215374	0
very good	0.5865014	0	0.8346413	0	-1.97679	0	-0.7913664	0
excellent	1.456079	0	1.405978	0	-2.556376	0	-0.8332158	0
const	-1.75361	0	-3.284437	0	-0.4734984	0	-3.416279	0

Hausman and Mcfadden Test for IIA: Professional is the excluded choice	
IIA Test: P-value	0.205
Log-likelihood	84.8
N	13653

Table 8:Results from the binomial logit estimations.

Variables	Non-participation		General		Vocational		Teaching/art	
	Coef	P-value	Coef	P-value	Coef	P-value	Coef	P-value
Parents' education:								
At least one parent: junior high school	0.1872426	0.102	0.3321928	0.008	-0.0290509	0.768	-0.1152526	0.343
Both parents:junior high school	0.1649364	0.116	0.3408315	0.002	-0.2968759	0.001	-0.1219914	0.277
At least one parent: high school	0.5445838	0	0.8744592	0	-0.3106484	0.004	0.0738801	0.563
Both parents: high school	0.9674779	0	1.394539	0	-0.6970522	0	0.0985578	0.507
At least one parent: university	1.728428	0	2.094455	0	-1.021009	0	0.4524498	0.104
Both parents: university	2.562305	0	2.661048	0	-0.9416951	0.163	0.417389	0.213
Father's occupation:								
Retailer	0.2408205	0.055	0.2545279	0.036	-0.6709934	0	0.0702996	0.65
Craft	0.1438922	0.319	-0.0095831	0.947	-0.2132325	0.101	-0.1096197	0.461
Farmer	0.1327528	0.528	-0.0820405	0.704	-0.0262298	0.88	-0.3032383	0.205
Entrepreneur	0.4844536	0.002	0.3203599	0.036	-0.8491609	0	-0.3245745	0.148
Professional	0.6250733	0	0.3611484	0.015	-0.7586202	0.001	-0.0414321	0.877
Manager	1.001828	0	0.8826189	0	-0.3678917	0.176	-0.2669332	0.39
Teacher	0.6476521	0.005	0.4784504	0.019	-0.3456049	0.365	0.5208624	0.094
White collar high level	0.109886	0.316	0.1925606	0.057	-0.403649	0.001	-0.3015364	0.032
White collar low level	0.14747	0.229	0.0710016	0.542	-0.3417412	0.01	-0.1608304	0.326
Blue collar high level	-0.2044273	0.038	-0.2794927	0.005	0.1191441	0.173	-0.1922012	0.08
Female	0.3283796	0	0.4699606	0	0.6807682	0	2.38602	0
Region of residence:								
Centre	0.0858379	0.322	0.1297423	0.107	0.0744456	0.46	0.3359236	0.003
North-east	-0.068851	0.467	-0.1607722	0.07	0.3455562	0.001	0.3081315	0.009
South	-0.1132975	0.175	-0.0372128	0.635	0.0674839	0.469	0.3002502	0.005
Year of birth:								
born in 1976	0.605601	0	0.3219738	0	0.2302364	0.001	-0.0788324	0.425
born after 1976	1.246652	0	0.8240649	0	-0.2626627	0.148	2.217853	0
Junior high school score								
good	0.1455471	0.083	0.5108301	0	-1.065649	0	-0.5677519	0
very good	0.5873023	0	1.291325	0	-1.953888	0	-0.8383732	0
excellent	1.441581	0	2.139757	0	-2.771275	0	-0.9130015	0
const	-1.672098	0	-2.868465	0	-0.5587571	0	-3.193265	0

Table 9 Covariate balancing indicators before and after matching for the different specifications.

Treatment	N1	Comparison	N0	Median Bias	Median Bias	% off support
	Before		Before	Before	After	
				(1)	(1)	
General	350	Professional	2580	16.03	2.7	0.0
	230	Vocational	1356	12.25	8.64	7.5
	244	Teaching/art	442	12.25	4.87	14.4
Professional	1415	General	359	7.19	4.31	39.6
	2346	Vocational	1356	7.19	2.18	5.9
	1861	Teaching/art	442	7.19	4.36	23.8
Vocational	694	General	359	14.27	4.12	38.6
	1302	Professional	2580	14.27	1.92	1.4
	1106	Teaching/art	442	14.27	3.7	13.9
Teaching/art	286	General	359	7.57	6.85	19.5
	360	Vocational	2580	7.57	3.85	2.7
	346	Professional	1356	7.57	4.8	5.3

with caliper=0.01

Treatment	N1	Comparison	N0	Median Bias	Median Bias	% off support
	Before		Before	Before	After	
				(1)	(1)	
General	284	Professional	2580	12.25	2.65	2.6
	109	Vocational	1356	12.25	5.59	14.6
	92	Teaching/art	442	12.25	7.82	33.3
Professional	569	General	359	7.19	3.03	68.4
	1488	Vocational	1356	7.19	1.32	27.7
	819	Teaching/art	442	7.19	3.58	58.3
Vocational	236	General	359	14.27	2.42	65.3
	1130	Professional	2580	14.27	1.52	5.7
	543	Teaching/art	442	14.27	3.8	45.2
Teaching/art	104	General	359	7.34	6.64	42.2
	277	Vocational	2580	7.57	1.96	5.5
	257	Professional	1356	7.57	3.89	10.3

with caliper=0.001

Table 9 (continuing) Covariate balancing indicators before and after matching for the different specifications.

Treatment	N1	Comparison	N0	Median Bias	Median Bias	% off support
	Before		Before	Before	After	
				-1	-1	
General	216	Professional	2580	12.25	4.96	23.1
	359	Vocational	1356	12.25	0	56.0
	56	Teaching/art	442	12.25	3.73	79.6
Professional	429	General	359	7.19	0	74.7
	1300	Vocational	1356	7.19	0.21	46.7
	719	Teaching/art	497	7.19	1.39	72.8
Vocational	178	General	359	14.27	0	69.5
	1356	Professional	2500	14.27	0.19	6.3
	400	Teaching/art	442	14.27	1.5	57.3
Teaching/art	64	General	359	7.57	3.7	60.4
	237	Vocational	2580	7.57	0.94	14.9
	216	Professional	1356	7.57	1.28	31.9
with caliper=0.0001						

Notes:

(1) Median absolute standardized bias before and after matching median taking all the regressors

$$B_{before}(X) = 100\frac{X_1 - X_0}{\sqrt{(V(X_1) + V(X_0)/2}} \qquad B_{after}(X) = 100\frac{X_{1M} - X_{0M}}{\sqrt{(V(X_1) + V(X_0)/2}}$$

Table 10: Results for the average treatment on the treated effects: OLS and PMS-ATT models (% earning gain).

	OLS(ATT=ATE)	PMS(ATT)	PMS(ATT)	PMS(ATT)
General vs Professional	**-18.68**	**-18.94**	**-17.39**	**-16.78**
General vs Vocational	**-14.85**	**-18.6**	**-18.97**	**-12.45**
General vs Teaching/art	-4.85	-7.5	5.08	8.59
Professional vs General	**18.68**	**16.73**	**10.96**	**9.26**
Professional vs Vocational	1.82	-0.008	-1.68	-1.78
Professional vs Teaching/art	**13.82**	**15.13**	**12.44**	**11.84**
Vocational vs General	**16.85**	**13.39**	**20.2**	**16.57**
Vocational vs Professional	-1.82	-3.52	-2.35	-1.35
Vocational vs Teaching/art	**11.99**	**7.03**	**10.36**	**10.09**
Teaching/art vs General	4.85	2.15	-10.61	-6.96
Teaching/art vs Professional	**-11.99**	-3.28	-0.39	0.09
Teaching/art vs Vocational	**-13.82**	-7.16	-7.58	-6.28
caliper matching		*0.01*	*0.001*	*0.0001*

Note: **Bold type** indicates statistical significance at 5% level. 95% bias-corrected standard errors obtained by bootstrapping.

Table 11: Selection-corrected Log Monthly Earnings Equations Estimates (using the procedure suggested by Lee). Standard errors are bootstrapped.

	General		Professional		Vocational		Teaching/art	
Variables	**Coef**	**P-value**	**Coef**	**P-value**	**Coef**	**P-value**	**Coef**	**P-value**
Parents' education:								
At least one parent: junior high school	-0.1603789	0.254	-0.032021	0.177	-0.0242752	0.372	-0.0084844	0.912
Both parents:junior high school	-0.0683874	0.584	0.0074001	0.7	-0.0173613	0.616	0.115583	0.072
At least one parent: high school	-0.1061358	0.458	-0.0359757	0.192	0.0542309	0.288	-0.0293259	0.719
Both parents: high school	-0.0588145	0.74	-0.0189401	0.634	0.1287206	0.172	0.0153207	0.882
At least one parent: university	-0.1901382	0.368	-0.1755344	0.045	0.0045385	0.973	0.0090804	0.956
Both parents: university	-0.089286	0.699	0.0224198	0.863	0.2092613	0.343	-0.3843357	0.452
Father's occupation:								
Retailer	0.0962718	0.489	0.068167	0.061	0.1763437	0.03	-0.0178764	0.849
Craft	0.1190191	0.346	-0.0477952	0.159	0.0794237	0.115	0.0712574	0.445
Farmer	0.1077409	0.752	0.1403466	0.001	0.0017906	0.972	-0.07236	0.718
Entrepreneur	0.3823655	0.013	0.0577917	0.159	0.1602455	0.176	0.1345745	0.329
Professional	0.0643591	0.681	0.0701329	0.141	0.2571119	0.04	0.00754	0.955
Manager	-0.0143406	0.942	-0.0162232	0.866	0.3189859	0.005	0.1411559	0.368
Teacher	-0.2106161	0.49	0.0320911	0.677	0.2847104	0.054	0.1648699	0.537
White collar high level	0.1602415	0.189	0.0574527	0.078	0.1016593	0.143	0.0722908	0.496
White collar low level	-0.0817613	0.602	0.1064456	0.001	0.0256724	0.745	0.2233185	0.009
Blue collar high level	0.008279	0.945	0.0751925	0.004	-0.0036044	0.89	0.0898796	0.114
Female	-0.2399849	0.019	-0.2607843	0	-0.2344971	0	-0.2621143	0.299
Region of residence:								
Centre	-0.1303307	0.164	-0.1218808	0	-0.0591086	0.032	-0.0583817	0.463
North-east	-0.0290458	0.765	0.0631997	0	0.06378	0.011	0.0701067	0.322
South	-0.1139957	0.301	-0.2777178	0	-0.293759	0	-0.3840701	0
Year of birth:								
born in 1976	0.0659016	0.512	-0.0214416	0.254	-0.0023042	0.934	-0.0097918	0.866
born after 1976	-0.0323631	0.806	-0.0747389	0.304	0.1766491	0.418	-0.2759887	0.166
High school score	-0.0115981	0.353	0.0002091	0.915	-0.0040209	0.334	-0.0015043	0.726
Junior high school score								
good	0.0748588	0.536	0.0500873	0.103	0.164336	0.116	0.0423494	0.483
very good	0.1565735	0.43	0.0345064	0.341	0.3449472	0.105	0.0270482	0.788
excellent	0.1728853	0.542	0.0408948	0.168	0.5300611	0.124	0.1206484	0.507
Correction term 1	-0.2754238	0.465	-	-	-	-	-	-
Correction term 2	-	-	-0.1306418	0.215	-	-	-	-
Correction term 3	-	-	-	-	0.3428361	0.125	-	-
Correction term 4	-	-	-	-	-	-	0.1467179	0.621
const	7.097532	0	7.200591	0	7.816757	0	7.540569	0

Table 11.1: Results for the average treatment on the treated effects: Lee model (% earning gain).

	Lee(ATT)
General vs Professional	**-23.01**
General vs Vocational	**-39.69**
General vs Teaching/art	**-7.50**
Professional vs General	**37.98**
Professional vs Vocational	**-19.63**
Professional vs Teaching/art	**-2.11**
Vocational vs General	**37.29**
Vocational vs Professional	**-2.84**
Vocational vs Teaching/art	**1.02**
Teaching/art vs General	**11.32**
Teaching/art vs Professional	**-17.84**
Teaching/art vs Vocational	**-18.79**

Note: **Bold type** indicates statistical significance at 5% level. Standard errors are bootstrapped.

Table 12: Separate analysis of women and men. Results for the average treatment on the treated effects using PMS-ATT model (%earning gain).

	MALES		
General vs Professional	-10.04	-10.94	-8.76
General vs Vocational	-3.04	-5.65	-7.82
General vs Teaching/art	-1.47	9.82	15.79
Professional vs General	2.87	0.81	0.49
Professional vs Vocational	0.44	-0.98	-2.06
Professional vs Teaching/art	**14.54**	**15.47**	**22.59**
Vocational vs General	-0.26	5.27	7.84
Vocational vs Professional	-7.2	-6.66	-5.98
Vocational vs Teaching/art	13.93	14.31	13.08
Teaching/art vs General	-13.91	-18.59	-24.53
Teaching/art vs Professional	**-10.81**	**-9.32**	**-14.42**
Teaching/art vs Vocational	-7.31	-10.11	-12.21
caliper matching	*0.01*	*0.001*	*0.0001*

	FEMALES		
General vs Professional	**-20.45**	**-18.57**	**-19.27**
General vs Vocational	**24.1**	**-22.57**	**-19.04**
General vs Teaching/art	-1.03	8.05	4.29
Professional vs General	**20.8**	**15.27**	**15.34**
Professional vs Vocational	-2.36	-2.2	-1.48
Professional vs Teaching/art	**11.97**	8.73	6.99
Vocational vs General	**15.75**	**27.12**	**25.61**
Vocational vs Professional	1.35	3.63	5.29
Vocational vs Teaching/art	6	**8.09**	**8.4**
Teaching/art vs General	-7.28	-7.97	-1.69
Teaching/art vs Professional	-0.94	2.96	3.36
Teaching/art vs Vocational	6.25	6.98	-5.26
caliper matching	*0.01*	*0.001*	*0.0001*

Note: **Bold type** indicates statistical significance at 5% level.

Table 13: Distribution of Participation and employment rates by high school types.

High school types	Participation to the labour market		Employment	
	no	yes	no	yes
Vocational	11.86	88.14	34.01	65.99
Professional	25.08	74.92	38.31	61.69
General	68.23	31.77	72.55	27.45
Teching/art school	29.08	70.92	64.29	35.71
Total	37.52	62.48	46.05	53.95

Table 13.1: Results for the average treatment on the treated effects using PMS-ATT model on: participation to the labour market, employment probability and earnings conditional on employment and participation.

	Participation		
General vs Professional	**-39.36**	**-39.89**	**-41.13**
General vs Vocational	**-39.49**	**-40.25**	**-40.09**
General vs Teaching/art	**-29.45**	**-27.65**	**-25.28**
Professional vs General	**23.89**	**25.58**	**24.75**
Professional vs Vocational	**-11.86**	**-11.6**	**-11.5**
Professional vs Teaching/art	3.81	3.99	2.16
Vocational vs General	**25.62**	**25.43**	**24.04**
Vocational vs Professional	-1.91	-2.61	-3.27
Vocational vs Teaching/art	**7.12**	**7.73**	**6.13**
Teaching/art vs General	**16.43**	**17.44**	**14.54**
Teaching/art vs Professional	**-12.15**	**-11.94**	**-13.07**
Teaching/art vs Vocational	**-17.51**	**-16.37**	**-17.28**
caliper matching	*0.01*	*0.001*	*0.0001*

	Employment		
General vs Professional	**-34.92**	**-37.5**	**-37.52**
General vs Vocational	**-31.72**	**-31.15**	**-31.5**
General vs Teaching/art	**-14.42**	**-14.79**	**-14.5**
Professional vs General	**16.38**	**18.47**	**17.06**
Professional vs Vocational	**-11.22**	**-13.69**	**-14.86**
Professional vs Teaching/art	13.15	**11.48**	**7.41**
Vocational vs General	**15.88**	**13.83**	**12.11**
Vocational vs Professional	-15.01	**-15.74**	**-15.56**
Vocational vs Teaching/art	9.01	**7.73**	**8.03**
Teaching/art vs General	5.96	9.41	6.04
Teaching/art vs Professional	**-23.24**	**-26.13**	**-27.59**
Teaching/art vs Vocational	**-24.73**	**-26.63**	**-28.38**
caliper matching	*0.01*	*0.001*	*0.0001*

	Earnings		
General vs Professional	**-12.56**	**-12.17**	**-14.16**
General vs Vocational	-10.1	**-14.63**	**-15.04**
General vs Teaching/art	1.7	3.88	12.12
Professional vs General	**10.83**	7.66	7.67
Professional vs Vocational	-1.32	-2.02	-2.14
Professional vs Teaching/art	**11.63**	**9.38**	**9.24**
Vocational vs General	**10.96**	**16.26**	**16.39**
Vocational vs Professional	-1.23	-1.44	-0.75
Vocational vs Teaching/art	**10.26**	**12.25**	**10.48**
Teaching/art vs General	4.24	4.16	-8.92
Teaching/art vs Professional	**-4.35**	-3.39	1.04
Teaching/art vs Vocational	**-8.02**	**-8.14**	-6.87
caliper matching	*0.01*	*0.001*	*0.0001*

Note: **Bold type** indicates statistical significance at 5% level.

Table 14: Results for the average treatment on the treated effects: OLS and PMS-ATT models (% earning gain) using the 2001 survey on the Italian high school graduates.

	OLS(ATT=ATE)	PMS(ATT)	PMS(ATT)	PMS(ATT)
General vs Professional	**-17.2**	**-14.96**	**-17.55**	**-18.8**
General vs Vocational	**-15.63**	**-15.57**	**-13.4**	**-12.81**
General vs Teaching/art	-5.76	-6.66	-5.88	-9.63
Professional vs General	**17.2**	**16.81**	**16.19**	**17.38**
Professional vs Vocational	1.56	1.55	0.56	1.73
Professional vs Teaching/art	**11.43**	**13.4**	**11.07**	**11.25**
Vocational vs General	**15.63**	**18.72**	**23.68**	**23.06**
Vocational vs Professional	-1.56	-1.08	-0.98	-0.5
Vocational vs Teaching/art	**9.86**	**10.31**	**7.51**	**7.04**
Teaching/art vs General	5.76	3.4	6.24	8.74
Teaching/art vs Professional	**-9.86**	**-8.77**	**-9.52**	**-8.95**
Teaching/art vs Vocational	**-11.43**	-8.03	-6.91	-7.19
caliper matching		*0.01*	*0.001*	*0.0001*

Note: **Bold type** indicates statistical significance at 5% level.

Appendix 1: summary statistics (1995 sample).

Variable	Obs	Mean	Std. Dev.	Min	Max
Log(monthly net wage)	5316	7.162169	0.4025213	5.521461	8.517193
School type:					
Vocational	18093	0.1906262	0.3928058	0	1
General	18093	0.2766816	0.44737	0	1
Teaching/art	18093	0.120986	0.32612	0	1
Professional	18093	0.4117062	0.4921561	0	1
Parents' education:					
Both parents: elementary school	17663	0.1665063	0.372545	0	1
At least one parent: junior high school	17663	0.1402933	0.3473009	0	1
Both parents:junior high school	17663	0.2155353	0.4112048	0	1
At least one parent: high school	17663	0.2090811	0.406664	0	1
Both parents: high school	17663	0.158297	0.3650296	0	1
At least one parent: university	17663	0.070826	0.2565413	0	1
Both parents: university	17663	0.039461	0.1946946	0	1
Father's occupation:					
Retailer	18093	0.0693638	0.2540789	0	1
Craft	18093	0.0867739	0.2815112	0	1
Farmer	18093	0.0324988	0.1773255	0	1
Entrepreneur	18093	0.0382468	0.1917969	0	1
Professional	18093	0.0509589	0.21992	0	1
other independent	18093	0.0117725	0.1078636	0	1
Manager	18093	0.0406787	0.1975503	0	1
Teacher	18093	0.0307301	0.1725903	0	1
White collar high level	18093	0.1538717	0.3608357	0	1
White collar low level	18093	0.0899243	0.2860812	0	1
Blue collar high level	18093	0.2272702	0.4190801	0	1
other dependent	18093	0.1465208	0.3536373	0	1
Female	18093	0.5640856	0.4958897	0	1
Region					
Centre	18093	0.2560659	0.4364707	0	1
South	18093	0.3425082	0.4745617	0	1
North-east	18093	0.1872547	0.3901267	0	1
Year of birth:					
Born before 1979	18093	0.3040402	0.4600125	0	1
Born in 1979	18093	0.5961422	0.4906832	0	1
Born after 1979	18093	0.0998176	0.2997649	0	1
Junior high school score					
Sufficient	16985	0.3493082	0.4767655	0	1
Good	16985	0.2827789	0.4503631	0	1
Very good	16985	0.1915219	0.3935104	0	1
Excellent	16985	0.1763909	0.3811636	0	1
High school score	18093	45.17261	6.910493	36	60
Grandfather with at least high school degree	18843	0.1055564	0.3072772	0	1
Number of siblings					
1 sibling	15525	0.591562	0.4915607	0	1
2 siblings	15525	0.2863124	0.4520517	0	1
3 siblings	15525	0.0820612	0.2744668	0	1
4 siblings	15525	0.0400644	0.1961166	0	1

Appendix 1: summary statistics (2001 sample).

Variable	Obs	Mean	Std. Dev.	Min	Max
Log(monthly net wage)	7686	6.686442	0.3969121	5.416101	8.987197
School type:					
General	18548	0.2009381	0.4007127	0	1
Vocational	18548	0.2828337	0.4503884	0	1
Teaching/art	18548	0.1116023	0.3148851	0	1
Professional	18548	0.4046258	0.4908327	0	1
Parents' education:					
Both parents: elementary school	18360	0.0825708	0.2752399	0	1
At least one parent: junior high school	18360	0.1184096	0.3231013	0	1
Both parents:junior high school	18360	0.2409041	0.4276439	0	1
At least one parent: high school	18360	0.2458061	0.4305758	0	1
Both parents: high school	18360	0.1884532	0.3910843	0	1
At least one parent: university	18360	0.0824619	0.2750746	0	1
Both parents: university	18360	0.0413943	0.1992059	0	1
Father's occupation:					
Entrepreneur	18114	0.0760738	0.2651234	0	1
Professional	18114	0.0500166	0.2179852	0	1
other independent	18114	0.1811858	0.3851827	0	1
Manager	18114	0.0446064	0.2064437	0	1
Teacher	18114	0.0640389	0.2448289	0	1
White collar high level	18114	0.1419344	0.3489925	0	1
White collar low level	18114	0.1004196	0.3005669	0	1
Blue collar high level	18114	0.3417246	0.4743009	0	1
Female	18548	0.550248	0.4974821	0	1
Region					
Centre	18548	0.2557149	0.4362741	0	1
North-east	18548	0.1820681	0.3859111	0	1
South	18548	0.3553483	0.4786316	0	1
North-east	18548	0.2068687	0.4050714	0	1
Year of birth:					
Born before 1979	18084	0.2945698	0.4558617	0	1
Born in 1979	18084	0.6635147	0.4725201	0	1
Born after 1979	18084	0.0419155	0.2004016	0	1
votoma	18360	2.221405	1.122998	1	4
Junior high school score					
Sufficient	18360	0.3283224	0.4696156	0	1
Good	18360	0.3257081	0.4686516	0	1
Very good	18360	0.1899237	0.3922513	0	1
Excellent	18360	0.1560458	0.3629086	0	1

Appendix 3: Matching algorithm

The matching algorithm applied in this paper is similar to Lechner (1999b). Thus, for the estimators and their covariance matrixes, the reader is asked to turn to that paper.

Matching algorithm 1. Specify and estimate a multinomial discrete choice model to obtain the (estimated) propensities P(T=1| X), P(T=2| X), P(T=3|X) and P(T=4|X). Test for omitted variables in a binomial framework. Compute the conditional probabilities $P^{m\|ml}(X)$. 2. Common support. Eliminate all observations outside common support. 3. Apply the following procedure to match each observation in group T=m with an observation in the comparison group T=l: (i) Choose an observation from the group m, and remove it from the pool. (ii) Find an observation in the group l that is as close as possible to the one collected in step (i) in terms of predicted probabilities. The distance can be measured by a Mahalanobis distance metric. Alternatively, base the closeness on the conditional probability $P^{m\|ml}(X)$. Do not remove that observation so that it can be used again. (iii) Repeat (i) and (ii) so that there is non observation left in the group m. (iv) Repeat (i)-(iii) for all combinations of m and l. 4. Test for balance of the covariates. In case that the covariates are not balanced, refine the specification of the discrete choice model, and go through steps 2-4 again. 5. Use the comparison groups formed in 4(iv) to compute the respective conditional expectations by sample mean. Note that the same observation may appear several times in the sample. 6. Compute the estimates of the treatment effects using results of step 6, and compute their covariance matrix.

Appendix 4: The proposed Italian high school system

The nowadays system consists of four types of high schools: licei (General high schools), istituti tecnici (Professional high schools), istituti professionali (Vocational high schools) and finally altri istituti (Teaching and Art schools).

The proposed system (statutory law n.53 28/03/2003) comprises only two types of tracks: the general high schools (*licei*) and the vocational track (*sistema dell' istruzione professionale*). It is worth stressing that it is always possible to bounce from one school type to the other attending some additional lessons (*passerelle*) provided either by the leaving or the arrival school.

The general high school system split up into: Liceo artistico; Liceo classico; Liceo economico; Liceo linguistico; Liceo musicale; Liceo scientifico; Liceo tecnologico; Liceo delle scienze umane. The general high school lasts five years: the first two years represent the first "biennio", the third and the fourth are the second "biennio"; it follows the fifth year with the final graduation (*esame di stato*).

The length of the vocational tack depends on the choices of the student and it is not defined a-priori. It is established a system of "*alternanza scuola-lavoro*" which enhances some work experiences organized by the vocational schools themselves and valued as formative activities like the didactic ones, putting schooling and working on the very same level. These "training" experiences are aimed only at students older than 15 years after some periods of formal education at school. At the end of the third year, the student takes a degree of "*qualifica*". If the student would rather not to go on to higher education he has the possibility to attend the fourth year taking the degree of "*qualifica quadriennale*". In the case the student prefer to enrol at university he has to attend the fifth year and take the final degree which has the same value of the general school one.

Figure 1: The dependent variable (log earnings; 1995).

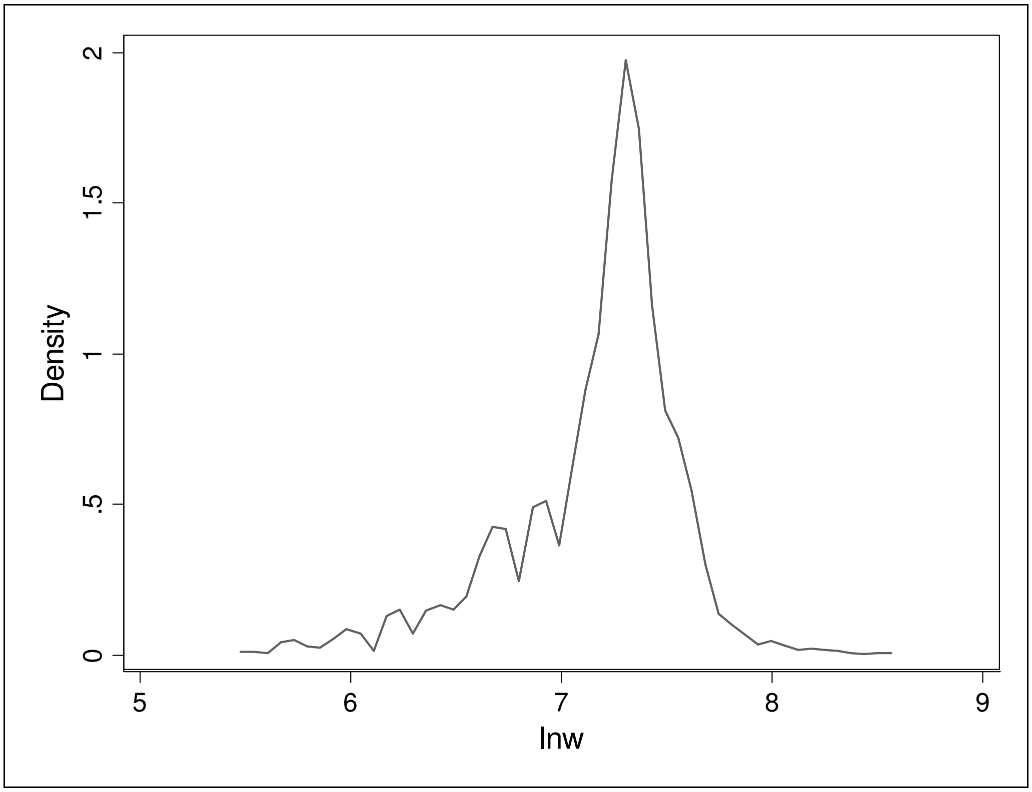


Figure 2: The dependent variable (log earnings; 2001 sample)

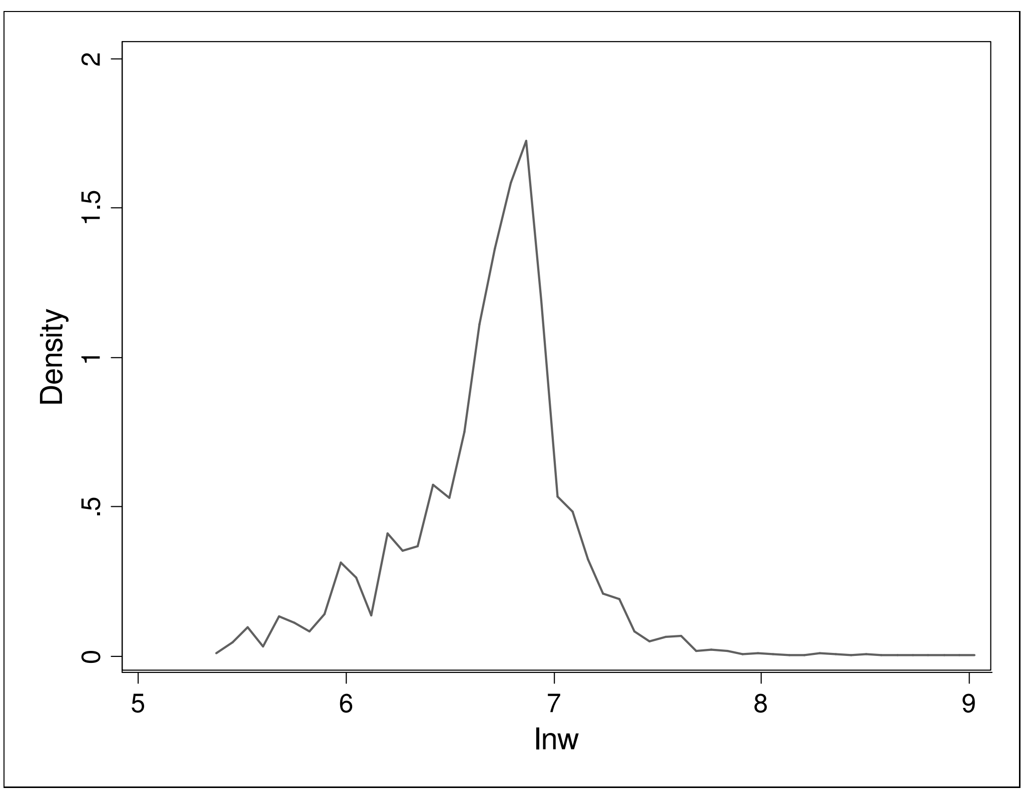
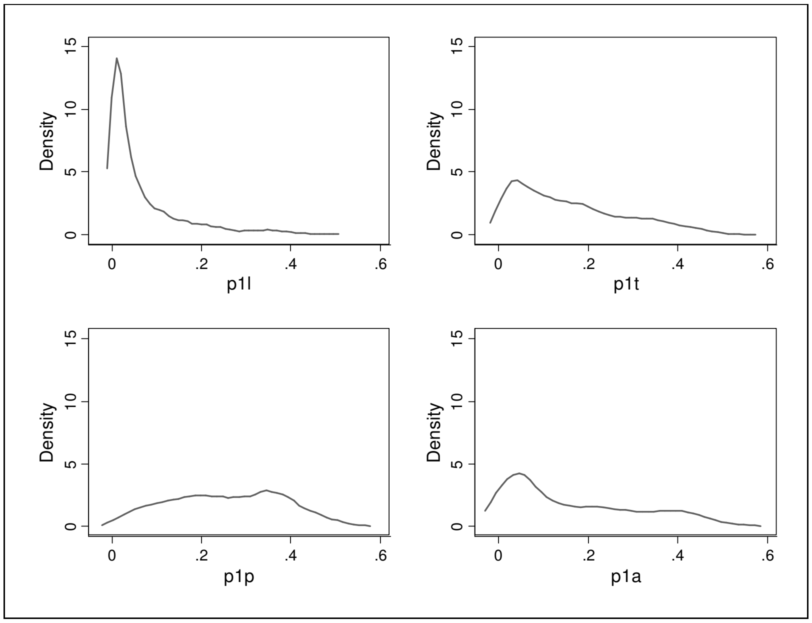
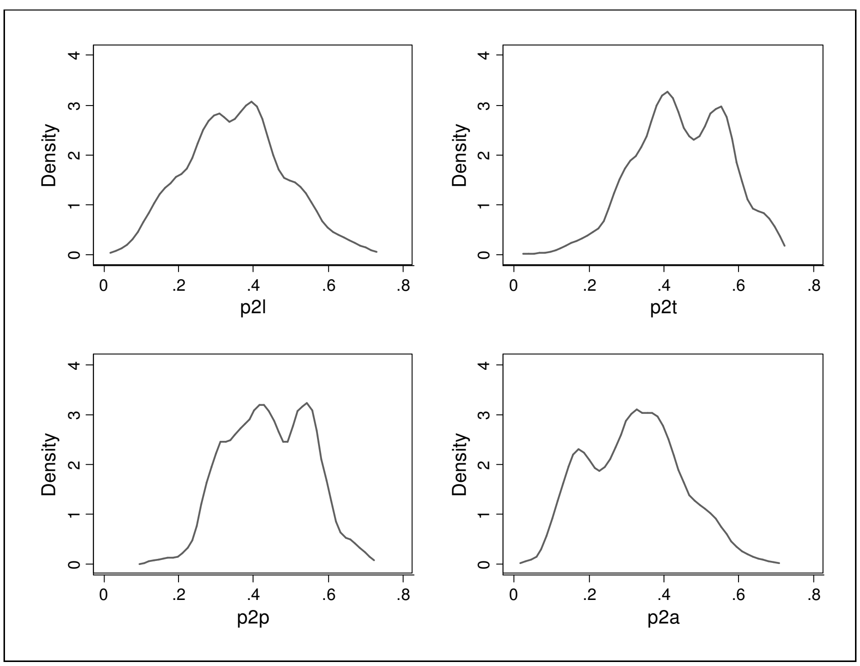


Figure 3: Distributions of the estimated propensities to be assigned into the four school types (1995 sample).

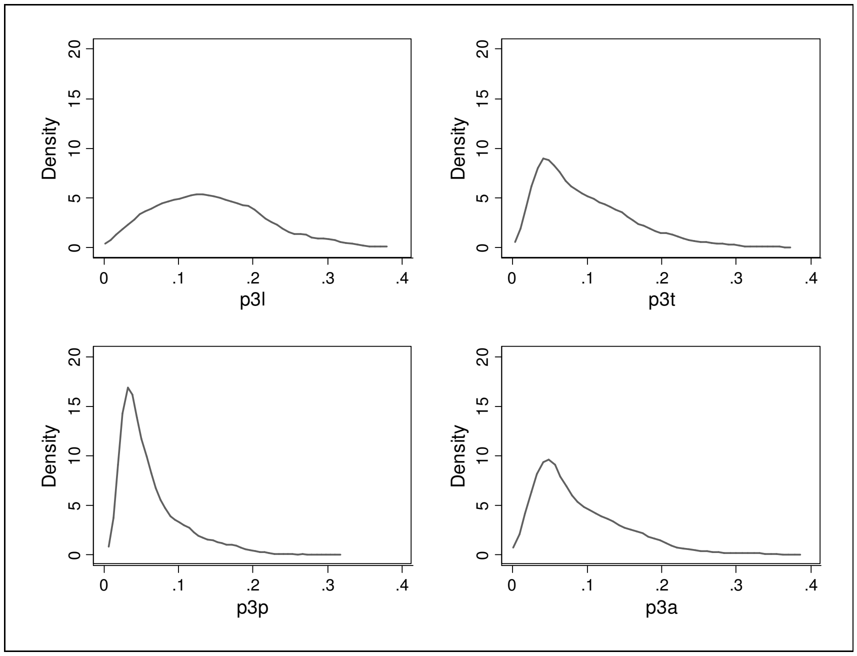
Estimates for propensity to enrol at Vocational high school.



Estimates for propensity to enrol at Professional high school.



Estimates for propensity to enrol at General high school.



Estimates for propensity to enrol at Teaching/art high school.

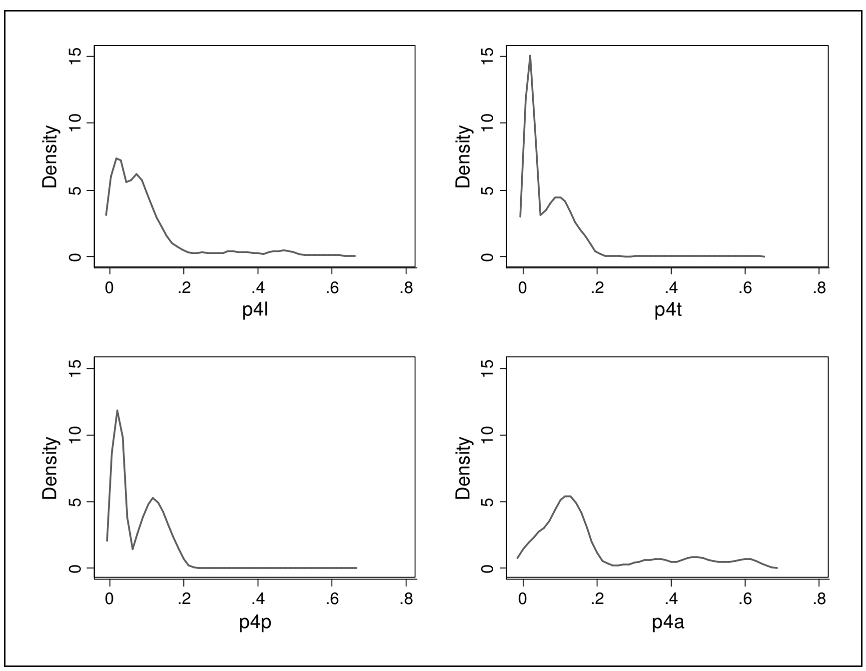
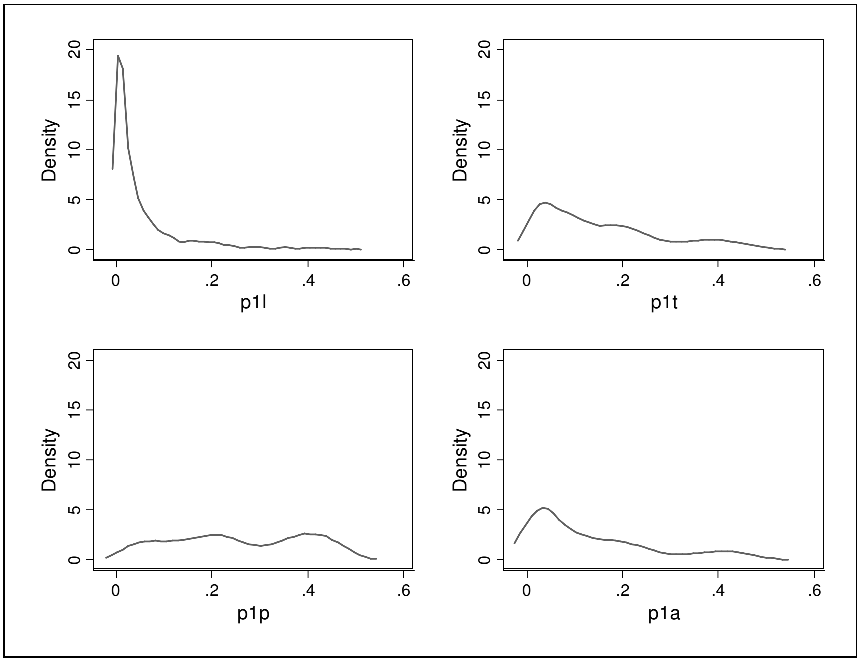
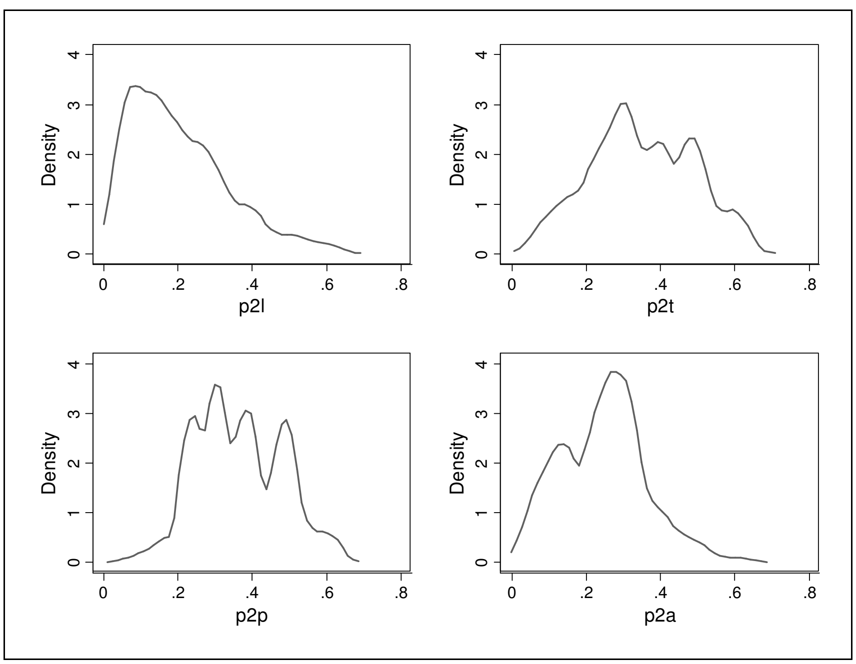


Figure 4: Distributions of the estimated propensities to be assigned into the four school types (2001 sample).

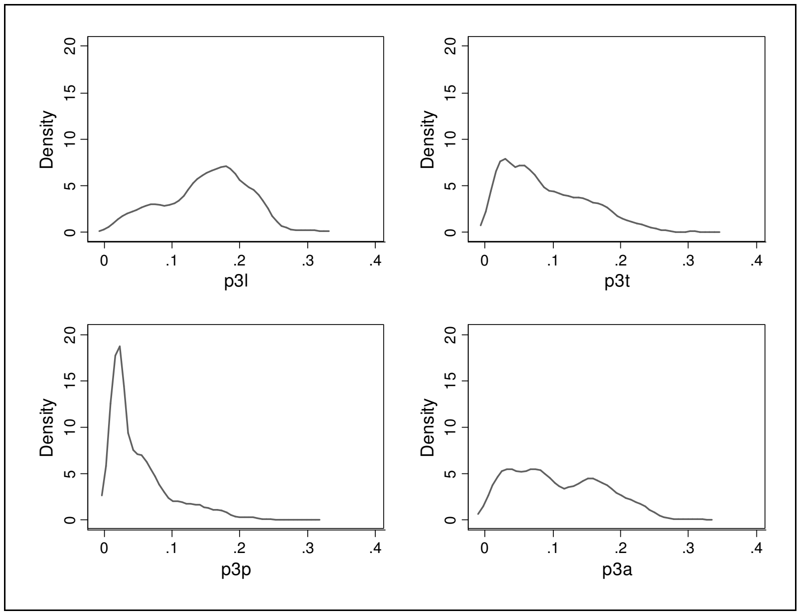
Estimates for propensity to enrol at Vocational high school.



Estimates for propensity to enrol at Professional high school.



Estimates for propensity to enrol at General high school.



Estimates for propensity to enrol at Teaching/art high school.