

Estimation of the University Education Effect on the Italian Graduates' Labour Income

Antonino Di Pino, Patrizia Pulejo⁽¹⁾

Dipartimento di Economia, Statistica, Matematica e Sociologia - Università di Messina.
dipino@unime.it, ppulejo@unime.it

Abstract: The aim of this work is to estimate, for a sample of Italian graduates surveyed in 2004, a wage equation in which the effects of individual ability and education can be estimated separately. The differential in individual wage of, respectively, a graduate and a holder of a high school diploma is estimated through a Propensity Score Matching procedure, where a selected sample of high school certificate holders drawn from the Bank of Italy *SHIW* Survey serves as a control group. Estimation results show that the average effect of college attendance on labour income three years after the degree is positive by about 16% (+24% for men and +8% for women). The gap between men and women seems larger in the South of Italy and in the Islands.

Keywords: Ability Bias, Propensity Score Matching, Misspecification, Endogeneity.

1. Misspecification and endogeneity in wage equation: methodological problems

Several problems occur in estimating causal effect of education on individual labour income. The *IV* approach suffers from identification problems of the schooling equation (Card, 2001, *inter alia*). An exclusion-restriction rule requires that at least one instrumental variable included in the education equation be left out of the income equation (labour income). Family background (parental education), proximity to colleges and modifications in the education law are examples of instruments that analysts generally introduce in the education equation⁽²⁾. The aim of this contribution is to improve the *IV* approach by adopting both *IV* and matching procedures to evaluate the influence of college attendance on labour income. To this purpose, we estimate the difference between the labour income earned by a graduate and a holder of a high-school diploma who are similar in terms of age, family size, status in the family and geographical area of residence. For the empirical analysis we select a survey sample from the Istat Survey on Italian Graduates in the year 2001, composed by 5536 individuals, of whom 4057 are employed (and 1479 are not). All the observed subjects are characterized by an identical level of schooling (degree). The control group, holders of high school diploma, is drawn from the Bank of Italy 2002 *SHIW* Survey. The

⁽¹⁾ This work has benefited from financial support of MIUR - National Research Project (PRIN 2008-2010)

⁽²⁾ Flabbi (1997), with the *IV* procedure, estimates the marginal effect of one year of education in Italy to be respectively 6,2% for men and 5,6% for women. Brunello and Miniaci (1999) show that, taking into account the endogeneity of schooling demand, the estimated return of an additional year of schooling increases from 4,8% to 5,6%.

methodological issues related to the utilization of a matching procedure as an estimation method are discussed in the next section.

2. Model specification and estimation results

The individual labor income function is specified as in Eq. (1), which allows for selectivity:

$$\ln W_i = \mathbf{x}'_i \boldsymbol{\alpha} + \mathbf{z}'_i \boldsymbol{\beta} + u_i \quad \text{if } L_i = 1; 0 \text{ otherwise} \quad (1)$$

$$L_i^* = \mathbf{y}'_i \boldsymbol{\phi} + v_i \quad \begin{cases} L_i = 1 & \text{if } L_i^* > 0 \\ L_i = 0 & \text{otherwise} \end{cases} \quad (2)$$

In Eq. (1), the dependent variable $\ln W_i$ is the logarithm of the monthly wage of a graduate. $\mathbf{x}'_i$ is a row vector that includes the proxy variables measuring individual ability, given by a dummy for “delay”, if the degree was obtained beyond regular completion time, and the difference in high school test score between graduates and high school certificate holders, $s_i - s_i^*$, where s_i is the high school test score of a graduate included in the Istat survey sample of Italian university graduates, and s_i^* is the corresponding high school test score of a matched individual with identical profile drawn by a control group of high school certificate holders, selected from the Bank of Italy *SHIW* dataset. Therefore, the variable s_i^* , obtained by adopting a matching procedure, serves to compute the instrumental variable $s_i - s_i^*$ utilized as a proxy of ability in the estimation of equation (1)⁽³⁾. $\mathbf{z}'_i$ is a row vector of exogenous variables (sex, part-full time job, area of residence, regional GNP *pro-capite*). The specification of the selection criterion is given by the participation equation (2), where the observed dependent variable, L_i , is a binary dummy variable equal to zero if the subject does not work. $\mathbf{y}'_i$ is a row vector of exogenous variables that includes among the regressors variables such as the month when the subject took his/her degree and parental education. Both these variables are assumed to be independent from $\ln W_i$ as required by the identification condition of the selection criterion function. The control group allows us, also, to assign the labour incomes of the high school certificate holders, W_i^* , to the “treated” (the graduates). In order to integrate the dataset of the graduates with the variables s_i^* and W_i^* taken from the control group, we adopt a Propensity Score Matching procedure, where the estimated probability of assignment of each unity is assumed to be conditional to a set of variables, H , including: sex (dummy), age, kind of high school (categorical variable), number of siblings, presence of children in the family (dummy), position of the subject in the family (categorical variable) and geographical area (dummy). We assume the variables conditioning the assignment, H , to be independent or, at least, exogenous with respect to the outcome variables, s_i^* and W_i^* ⁽⁴⁾. Furthermore, as identification condition, we impose that a common characteristic of the individuals included in the control group is to be employed or to be looking for a job in

⁽³⁾ s_i^* can be assumed as an instrumental variable exogenous with respect to the individual wage. In this context, matching can be considered as the first stage of a Two-Stage estimation procedure.

⁽⁴⁾ This condition in matching procedure ensures that assignment is not influenced by the outcome variables and, consequently, that matched observations do not self-select themselves.

the labour market. In this context, the causal effects may be formalized using the following notation:

$$E(\ln W_i - \ln W_i^* | H_i, L = 1) = E(\ln W_i) - E(\ln W_i^*) - E(u_i - u_i^* | H_i, L = 1) \quad (3)$$

Eq. (3) shows that the effect of the treatment (attendance of college) varies from person to person, as a consequence of interpersonal variation in H and of the possible presence of unobservable, confounding variables in both the error terms u_i and u_i^* ⁽⁵⁾. A two-stage Heckman's procedure is adopted to estimate $\ln \hat{W}_i$ from equations (1) and (2) (Heckman, 1979). What we obtain is a consistent estimation of labour income: $\ln \hat{W}_i = \mathbf{x}_i' \hat{\alpha} + \mathbf{z}_i' \hat{\beta}$, where $\mathbf{x}_i' \hat{\alpha}$ can be considered as the estimated influence of individual ability on labour income. Then we compute the differences $\ln \hat{d}_i = \ln \hat{W}_i - \mathbf{x}_i' \hat{\alpha} - \ln \hat{W}_i^*$ and we consider the differentials $\hat{d}_i$ as the estimated causal effects of college education on labour income corrected by individual ability influence.

Tab. 1 - Estimation of the Individual Labour Income Equation

	Estimator: Two-Stage Heckman (II stage: OLS regression)	Dependent variable: $\ln W = \log$ of individual monthly labour income		
	Explanatory variables:	coeffs	s.e.	p-value
$\hat{\beta}_0$	Intercept	6.3574	1.0599	0.0000
$\hat{\alpha}_1$	Dummy: degree obtained beyond regular completion time = 0; degree obtained at its regular end = 1)	0.0749	0.01249	0.0000
$\hat{\alpha}_2$	difference in high school test score between graduates and high school holders.(*)	0.0013	0.0009	0.0839
	heckman correction for selectivity	-0.0885	0.0522	0.0451
$\hat{\beta}_1$	Fixed effect: Regional Inequality Gini Coefficient	0.5937	0.2746	0.0153
$\hat{\beta}_2$	dummy: Sex (0= man; 1=Woman)	-0.1357	0.0226	0.0000
$\hat{\beta}_3$	dummy: Full-time (0= part-time job; 1= full-time job)	0.6035	0.1006	0.0000
$\hat{\beta}_4$	dummy: Area (South=0; North-Centre=1)	0.0822	0.0137	0.0000
(*) The scores of high school holders are drawn from SHIW dataset				
sample size: n = 4057; (missing data: 722)				
Adj-R ² = 0.3997 F (7; 3327) = 246.05 p-value = 0.0000				

⁽⁵⁾ We utilize a sensitivity test (Rosembaum T statistic) to verify the significativity of influence of "hidden" confounding variables on the treatment effects

Table 1 shows the estimated coefficients of the labour income equation (1). We use the estimated coefficients to compute the differentials $\hat{d}_i$ and the ratios $\hat{d}_i/\hat{W}_i$ for different profiles of graduate subjects. A few selected results are reported in Tab.2a and Tab.2b. We have to take into account that several potential unobservable factors can affect the robustness of estimates and, consequently, of the variability of the results of tables 2a and 2b. However, we may generally evaluate about 15% the impact of college education

Tab. 2a - Average college-education effect on individual income (in percent). Classification for geographical area and sex				Tab. 2b - Average college-education effect on individual income (in percent). Classification for cultural sector		
	$\hat{d}_i/\hat{W}_i$					
Geographical Area	Males	Females	All	Cultural Sector	$\hat{d}_i/\hat{W}_i$	one year of education average effect
North-West	27.53%	14.14%	20.66%	Humanistic Sciences	6.37%	1.03%
North-East	24.98%	6.98%	14.93%	Social Sciences and Economics	16.90%	2.64%
Centre	24.49%	10.05%	16.48%	Exact Sciences, Chemistry and Technology	17.43%	2.71%
South	19.66%	0.75%	11.44%	Jurisprudence	17.05%	2.66%
Islands	12.28%	-12.85%	-2.23%	Engineering and Architecture	22.33%	3.42%
Italy	24.00%	7.91%	15.63%	Medicine	8.51%	1.37%
				Physical Education	13.81%	2.18%
				Tot	15.63%	2.45%

on labour income of the subject three years after his/her degree. This effect seems to be lower for women, particularly in the South and the Islands. Engineering and Architecture should be the most remunerative educational faculties for Italian graduates.

References

- Bank of Italy (2004), *Survey on Household Income and Wealth of Italian Families in 2002*.
- Brunello G. - Miniaci R. (1999), The Economic Returns to Schooling for Italian Men. An Evaluation Based on Instrumental Variables, *Labour Economics*, vol. 6, n.4, 1999, pp. 509-19.
- Card D. (2001), Estimating the return to schooling: progress on some persistent econometric problems, *Econometrica*, 69: 1127-1160.
- Flabbi L. (1997), Investire in istruzione: è meglio per lui o per lei? Stima per genere dei rendimenti dell'istruzione in Italia, Milano, *Università degli studi di Milano - Bicocca, Working Paper*, n. 08, 1997.
- Heckman J. J. (1979) Sample selection bias as a specification error, *Econometrica*, 47: 153-161
- Istat – Istituto Nazionale di Statistica (2004), *Survey on Italian Graduates in 2001*.
- Rosenbaum P. R., Rubin D. B. (1985) Constructing a control group using multivariate matched sampling methods that incorporate the propensity score, *American Statistician*, 39: 33-38
- Rosenbaum P.R. (2002) *Observational Studies*, 2d ed., Springer, New York