

Selectivity and Heterogeneous Returns to Schooling in Italy: An Evaluation Based on Control Functions

PAOLO GHINETTI

paolo.ghinetti@eco.unipmn.it

Università degli Studi del Piemonte Orientale

INTRODUCTION

The estimation of economic average returns to education is subject to a number of well-known econometric problems that affect standard OLS estimates. Beside measurement errors, the first is the so-called 'ability bias' which is responsible for the 'endogeneity' of education in earnings equations. This bias is caused by the fact that the intercept is individual-specific and unobservable (ability), i.e. unobservable characteristics correlated with education that make individuals more productive and earning more whatever is their education level. The second is the so-called 'return bias', due to heterogeneous returns to education which make individuals selecting into given education levels according to specific unobservable gains.

If this is the case, non only the intercept, but also the slope of schooling in the earnings function is individual-specific.

In many empirical applications, instrumental variables (IV) are used to tackle selectivity issues. The idea behind IV is simple: if both observed (education) and unobserved ability affect individual earnings, to the extent which the two measures of ability are correlated standard estimates of average returns to education are biased. This 'ability' bias can be avoided if at least one valid instrument – correlated with ability, uncorrelated with earnings - is available. A well-

known limitation of IV is that they provide consistent estimates of average returns to schooling only if they are homogeneous in the population, and subject only to endogeneity and the associated 'ability bias'.

If returns to education are heterogeneous, estimates by instrumental variables are able to identify only 'local' effects (LATE), i.e. the average return for the subpopulation of individuals affected by the instrument. This makes OLS and IV estimates not directly comparable, as they refer to different populations.

The aim of this paper is to overcome these limitations and to use a simple control functions (CF) estimator proposed by Card (2001) to get consistent returns to schooling in a model which allows for both heterogeneous ability and heterogeneous returns to education. The CF approach is more flexible than IV as correction terms included in the earnings equation simultaneously account for both endogeneity and self-selection, to obtain estimates of returns to schooling robust to both the 'ability' and the 'return' bias.

We estimate this model using data for Italian men in the 25-60 age interval. The sample is drawn by the seven most recent waves (1995 to 2010) of the Survey on Household Income and Wealth (SHIW) run every two years by the Bank of Italy. Similarly to other papers on education returns in Italy, for the identification of schooling we take advantage of a major reform which in 1962 increased the number of years of compulsory schooling, and use a set of cohort dummies as exclusion restrictions.

The paper is organised as follows. Section 2 discusses a number of methodological and econometric issues. Section 3 reviews the empirical literature for Italy. Section 4 describes the main features of the data. Section 5 discusses estimation techniques and the identification strategy. Section 6 presents the main econometric results. Conclusions follow in Section 7.

METHODOLOGICAL AND EMPIRICAL ISSUES

If individuals are heterogeneous with respect to characteristics that affect both economic choices and wages, they may not be allocated randomly into sectors and education levels. This introduces potential sources of bias in the least square estimation of the average return to education (or to different schooling levels) in the population.

First we have the 'omitted variable' or 'selection' bias which occurs when returns to schooling are homogeneous, but more educated workers have on average different unobservable productivity than otherwise similar less educated individuals.¹ If this is the case, more inherently productive individuals would earn more (or less, if the correlation is negative) at any schooling level. From the empirical viewpoint, the intercept of the wage function is individual specific (see Blundell et al., 2005; Card, 1999). This correlation is in general considered positive ('ability bias'), inducing an upward bias in the OLS estimation of the returns to education.

A second source of bias arises when returns to education are heterogeneous and individual specific. On the empirical side, the error of the wage equation includes a term that captures

¹ For example, this happens if, conditional on the set of X controls included in a standard log wage equation, earnings are correlated with individual traits that reduce the cost of acquiring higher education.

the comparative wage advantage due to unobservables of individuals in acquiring specific education levels. If the comparative advantage differs by schooling levels, it is correlated with the corresponding endogenous variable. If people self-select into sectors or education levels based on unobservable gains, OLS estimates suffer from a 'return bias'.

The direction of the bias is not obvious. In the more plausible scenario, people choose the level of schooling that gives the highest wage gain. Selection is therefore positive and OLS log wage estimates are upward biased. Negative selection would arise if, instead, jobs requiring specific education levels are on average associated with better non monetary attributes, and therefore chosen even in presence of lower wages. If this difference is positive (negative), workers are efficiently allocated in the education group in which their productivity is higher (lower).

The third potential source of bias in OLS estimates is measurement error. If errors are not correlated with education, returns estimated by OLS are downward biased.

The main strategy used in the economics of education literature to retrieve causal schooling effects is to estimate mincerian wage equations by instrumental variables (IV). Although it does not require any functional or distributional assumption about the stochastic relationship between unobservable determinants of earnings and schooling choices, the IV estimator has the well-known limitation that it is able to recover average causal effects only if they are homogeneous (endogeneity case), and in case of measurement errors in the variables of interest (Card, 1999; Blundell et al, 2005).

When returns are heterogeneous ('return bias'), OLS and IV estimates are not directly comparable because the latter are an unbiased estimate of the weighted average of the effect for the subpopulation of individuals affected by the instrument(s), and not of the average return in the whole population.

Instead of trying to eliminate the bias as IV do, the CF estimator controls for it. In general a CF approach is implemented with a two-step procedure, in which the residuals from the first step reduced form selection equation(s) for the endogenous variable(s) are added to the wage equation in the second step (Card, 2001; Blundell et al., 2005). It requires specific assumption about the type of bias that needs to be corrected, for example whether, conditional on the X , the wage error term contains only an individual intercept or also individual gains from unobservables; second, by making assumptions about the stochastic relationship between unobserved wage components and the endogenous variable(s), in particular on the functional form taken by the mean of the error of the outcome equation conditional to the endogenous variables. The CF estimator is in general implemented by assuming that these conditional means are linear functions of the errors of the selection equation for schooling, known up to a number of estimable parameter.

Once these conditional means are estimated and included in the set of regressors of the outcome equation, OLS on the selectivity-corrected wage equation are consistent and free of bias of the specified form. In presence of endogeneity (random intercepts) the control function is the overall residual of the selection equation. In case of random slopes the correction term is the residual time the schooling variable. Since the coefficients of the correction terms are proportional to the covariances between the endogenous processes and the two components of

the outcome's error term, the CF approach provides consistent estimates of average returns also in case of self-selection and allow to test directly for the direction and the intensity of endogeneity and self selection (if any). One pitfall is that with the general CF approach it is difficult to account for observed heterogeneity in returns to education: interactions between, schooling and other X variables result in additional endogenous variables (the original one plus the interactions), which would require a sufficient number of instruments to predict all the interactions properly. Unless making distributional assumptions such as normality in the Heckman model, IV and CF methodologies both suffer from this limitation.

Traditionally, the main variables used as instruments for schooling are policy reforms such as changes in compulsory schooling years, or family background variables, such as parents' education or employment sector.

According to Card (1999), family characteristics such as father's and mother's education are not legitimate instruments for observed education. Whilst the transmission of the social and economic status partly occurs through the accumulation and development of human capital, skills and abilities and, therefore, before entering the labour market, even after controlling for the occupational and the schooling level of the sons, family characteristics may directly influence individual wages, so that controls for them should be included in wage regressions².

Due to the intergenerational correlation between parents' and sons' ability, an IV approach based on father's and mother's schooling levels may be even more upward-biased than OLS (Card, 1999). The distortion is proportional to the intergenerational correlation in abilities and the effect of parent' ability on their own education level. If the latter is small, the distortion that the former imposes on the IV estimates is small. While it is possible to evaluate the goodness of family background instruments with tests on the mis-specification of the IV model such as the Sargan test on over-identifying restrictions, as a general empirical rule Card (1999) suggests that key family characteristics such as parents' education should be added as additional controls to the wage equation. Their inclusion reduces the upward bias in OLS estimates of returns to education.

By converse, policy reforms such as changes in compulsory schooling or that modify the access to particular curricula are typically exogenous. By reducing the marginal cost of schooling for students in specific cohorts without any residual effect on unobservable earnings determinants, they are considered legitimate instruments. However, using differences between individuals who simply attended school or worked in the public sector in different periods to draw causal inference requires particular care. The key assumption is that individuals in the pre and post reform periods are otherwise equal, i.e. that the distribution of unobserved determinants of outcomes and endogenous variables does not change over time.

² For example, sons of highly educated or highly qualified parents may be advantaged when searching for a new job as they can use better informal channels to access jobs which pay more, implying a causal structural effect of parental background on earnings. Even without an independent effect on wages, background variables are likely to be correlated with individual unobserved wage determinants. Indeed, they may proxy for unobserved individual skills and ability, especially when the endowment of human capital does not perfectly map into education levels and it positively depends also on unobserved skills and abilities inherited or learned from the parents.

This may not be case even if the distribution of abilities, returns and tastes is time invariant. First, wages can change over time for unobservable reasons such as technology shocks not correlated with schooling. Similarly, and perhaps more importantly, different cohorts may be more or less likely to study more for reasons unrelated to the policy reform, such as a secular trend in schooling that produces by itself inter-cohort growth in educational attainment. Allowing for these systematic differences is of course important when the identification strategy pools many cohorts of individuals.

RELATED LITERATURE

A number of empirical studies analysed returns to education in Italy by comparing alternative estimators such as OLS and IV and using data from different waves of Bank of Italy's SHIW (Survey of Household Income and Wealth). The educational structure is summarised by imputed years of schooling based on the highest attained degree. In general, this measure of years of schooling is rather inaccurate and potentially subject to substantial measurement errors: Given that many students face repetitions and failures, the expected duration of a degree for many individuals is higher than the legal one. Moreover, for many college students the enrolment can continue for years after the prescribed duration of the curricula. Finally, many students may have spent time in education before dropping out school without completing a degree.

Early studies used family characteristics (typically, mother's, father's and spouse age and education) as instruments for years of education. Using a sample of household heads full-time male workers in the 22-60 age interval and controls for age and its square, Cannari and D'Alessio (1995) find that IV estimates (6.08%) are considerably above OLS estimates (5.03%). Colussi (1997) uses a similar set of instruments and one additional dummy for individuals who went to schools in the 1942-48 period, aimed at capturing exogenous variations in education induced by the second world war, and find similar results.

More recent works have used schooling reforms as sources of exogenous variation in education. Flabbi (1997) analyses yearly earnings with SHIW data for 1991.³ Two institutional changes are used as instruments. The first is the presence of a university in the province of birth/residence of workers when they were 19 years old. The second is the 1962 law secondary school reform, which since 1963 raised the years of compulsory schooling from 5 to 8.⁴ The students affected are identified by a cohort dummy for people born since 1951.⁵ The analysis

³ Since at that time months worked within the year were not available, it was not possible to compute an estimate of hourly wages. However, the 1991 wave is the only cross section in which the information on birth provinces is public and not confidential, and this is key for the identification strategy.

⁴ Until 1963, only primary school was compulsory for children aged 6 to 11. Since October 1 1963, the leaving school age is 14, thus adding three further years of compulsory low secondary school. If the individual incurs in failures and repetitions, he or she is allowed to drop out at 14, even without finishing the junior high school.

⁵ Approximately the same individuals were also interested by a small-scale reform which in December 1969 extended the possibility of university enrolment to individuals with a five-years degree of secondary education, independently of the track (general or vocational). Before 1969, only individuals with a degree from high-schools (general education) could have direct access to university. Graduates from technical high schools had to undertake an exam for being admitted to college. The reform abolished the admission exam. Since the expected age of secondary school completion is 18-19 years, this opportunity was mainly available for people born since 1951.

of Flabbi (1997) is limited to the cohorts of individuals born between 1946 and 1962. Otherwise the 1951 cohort dummy may capture other things than the effect of the policy reform. For example, on average, older individuals are more educated because many younger individuals are still studying and therefore not in the sample of working people; second, exits from the labour market occurs on average at older ages for the more educated. Using a rich set of controls, males have higher IV returns to schooling (5.2%) than females (3.6%), the opposite of what happens with OLS (1.8% and 2.5%, respectively).

Brunello and Miniaci (1999) combine family background variables with the 1951 cohort dummy as instruments. Their sample consists of male household heads full-time/full-year workers aged 30 to 53 in the 1993 and the 1995 SHIW. By controlling for age, region and municipalities, IV estimates of returns to schooling are 5.7%, higher than the OLS value by about 20 percent (4.8%). They also estimate an empirical specification which includes a set of dummies for the highest completed level of education, thus allowing marginal returns to vary with schooling levels. IV returns from an additional year of schooling are approximately equal to 5% in junior high school, 4.2% in upper secondary and 7.2% in university, significantly higher than OLS returns.

Brunello (2002) uses a specific question included in the 1995 SHIW to construct a measure of individual's absolute risk aversion, which is then used as an instrument for education. Returns to education measured with IV are much higher (7.8%) than corresponding OLS ones (4.7%), almost by a 65%

The institutional features of the 1962 reform have been analysed in more detail by Brandolini and Cipollone (2002). They conclude that the reform was mostly effective for the cohort of individuals born in 1951 and 1952, who finished the primary school by 1962 and 1963. The reform was partly effective also for individuals born from 1949 to 1951, who were less than 15 years old by the time the reform was introduced. Finally, the reform was potentially effective also for individuals born between 1953 and 1956, who in 1963 were attending primary school. Accordingly, instead of relying on a single dummy for post 1951 cohorts, their exclusion restriction for education is a dummy for cohorts in the 1949-56 interval. Estimates of returns to education based on a sample of 30-60 years old full-time females in the pooled 1991-98 SHIW are in the range from 7 to 10%, depending on the method and the specification used.

Overall, the above studies shows that IV estimates of the return from one additional year of schooling are systematically higher than OLS ones by 19-23%, too large for being motivated only by measurement errors in education, which, according to Card (1999) should account for no more than a 10% gap. One possibility is that abler individuals have better job opportunities at any school level, so that their opportunity cost of staying in education is higher which induces them to drop out earlier and start working (negative 'ability bias'). The alternative explanation, is that the marginal returns of individuals affected by an exogenous increase in education

Since the instrument is a cohort dummy, first, older individuals are on average more educated because in younger cohorts many individuals are still studying and therefore are not in the sample of workers. Indeed, if an equation for schooling is estimated on the full sample of working individuals the coefficients for the age regressor is in general positive. Second, exits from the labour market occurs on average at older ages for the more educated.

induced by the instruments – especially school reforms and higher risk aversion – are higher than the average. If the marginal productivity of education is decreasing, the returns from additional schooling of individuals affected by the reform are higher than that of a random individual, and IV or CF that do not control for comparative advantages would overestimate the returns in the whole population. If heterogeneity matters, it should be accounted for as much as possible. Since one obvious source of this heterogeneity is parental background, Card (1999) suggests that more reliable IV or CF estimates can be obtained when controls for parental education are added to the wage model.

DATA AND DESCRIPTIVE STATISTICS

We use data drawn from the 1995, 1998, 2000, 2002, 2004, 2006, 2008 and 2010 waves of the Bank of Italy SHIW. In any year the survey covered of approximately 8,000 households, corresponding to around 21,000 individuals and 14,000 labour income earners.

The dataset is a cross section, but a panel component has been added since 1987. The size of the panel component increased over the years and in recent waves represent half of the overall sample (about 4,000 observations in 2002). In the present study, we use a pooled cross-section for the years of interest. Pooling is useful both to increase the sample size and to get estimates which are not sensitive to the choice of a specific year. Because of the panel component, the same individual may appear in different years. We eliminate “double-counting” by dropping all the observations for the same individual except the first one, i.e. when the household was randomly selected from the population.

The construction of the sample used in the empirical investigation follows the criteria used by many studies reviewed in Section 3. We restrict our sample to male household heads (approximately 2,600 in 2002) to avoid issues of female labour force participation and household formation.⁶ We further restrict the analysis to full-time employees who work in the non-agricultural sector (937 observations), who represents the typical male employee in the private and public sectors. We limit the analysis to workers in the age interval 25-60. These selection criteria reduce the 2002 sample to about 900 units. Older individuals have been excluded to avoid the problem of endogenous retirement, which is problematic especially for more educated and public employees. Especially in the recent past, the latter benefited from more favorable terms and conditions as compared to workers in the private sector.⁷

The final sample has been obtained also excluding observations with missing values for family background variables. Given these selection criteria, the 2002 final sample contains about 700 valid observations, which become approximately 5,300 for the pooled 1995 – 2008 sample.

⁶ Main categories excluded: retired people, unemployed, self-employed and students.

⁷ Of course, in the 25 to 30 age interval some individuals may be still enrolled in education and therefore not working, so that the probability to observe any given individual in the sample (of workers) is not independent to his schooling level. We experimented a bit with different age thresholds, for example by limiting the sample to the 30-55 age interval. Results are very similar in the two cases. We use the less restrictive definition of the age interval to work with a higher number of observations.

The SHIW provides a measure of annual earnings inclusive of extra-time compensations and fringe benefits, and net of taxes and social security contributions.⁸ Additional information is on the average number of hours worked per week and on the number of months worked per year. Based on that, we follow most empirical studies and construct an estimate of hourly net wages (inclusive of fringe benefits), which is obtained dividing annual earnings by months worked plus number of average weekly hours plus 52/12 (which is an estimate of the number of weeks worked per month). In order to analyse the robustness of results to the choice of the earnings variable, we also analysed monthly earnings and find results in line with those obtained with hourly wages.⁹ Nominal wages have been transformed into real ones (2008€) using the consumer price index.

In the Survey, the educational structure is summarised by a set of dummies for the highest completed schooling level. We use this information to define two education variables. First, we proceed similarly to many existing studies for Italy and impute years of schooling using the normal length of attained degrees (0 = no degree; 5 = primary school; 8 = junior high school; 11 = vocational schools; 13 = general high school; 16 = short university diploma; 18 = standard university diploma, laurea; 20 = post graduate diploma) to get a continuous variable for education. Second, we construct an ordered variable, which in principle can go from 0 to 7, increasing in the level of education.

The Survey provides also detailed information on the socio-economic family background, like parents' level of education, occupation, sector of employment. One disadvantage is that household heads are asked to recall the above information when his parents had his current age. Beside measurement issues, it is not obvious whether this information is the most relevant. We use dummies for the highest level of education of the mother and of the father, and for the occupation of the father. We also experimented with a dummy for the mother not employed but since it was never significant we excluded it from the final specification.

Additional information available in the Survey and used in the empirical analysis includes labour experience, age, geographical area of residence, size of the municipality of residence and time dummies.

A description of the sample and of the variables used in the empirical analysis and summary statistics are in Table 1.

The average of imputed years of education is about 11, which corresponds to a three years vocational degree in secondary schools. Consistently, mean of the ordered variable for the level of education is roughly 4, which corresponds to the same schooling as years of education.

⁸ We also experimented a bit by estimating equations for log wages without fringe benefits and non-monetary compensations, but the results were basically the same as those reported in Section 6.

⁹ In the context of public-private wage differences, monthly earnings are affected by potentially large differences in hours worked across sectors. It is possible to control for hours worked by including them among the regressors, but they would be endogenous to wages.

Table 1 – Variables' description and means

Variable	Mean	Std. Dev.
Real Hourly wage	10.286	6.341
Experience	23.881	9.578
Years of Schooling	10.911	3.829
School level	4.171	1.486
1=no degree	0.64	
2=primary	9.16	
3=junior high	34.24	
4=vocational	9.33	
5=high school	34.01	
6=short univ	0.88	
7=university	11.19	
8=postgrad.	0.56	
Area of residence: North-West	26.51	
Area: North-East	21.84	
Area: Centre	20.91	
Area: South	21.41	
Area: Islands	9.33	
Size of municipality<20,000	25.77	
20,000<size<400,000	62.84	
size>400,000	11.39	
Cohorts pre 1949	16.38	
Cohorts 1949 or 1950	6.75	
Cohorts 1951 or 1952	6	
Cohorts 1953 to 1956	14.01	
Cohorts post 1956	56.86	
Father no degree	17.28	
Father: Primary	51.29	
Father: Junior high school	19.32	
Father: Secondary high school	9.15	
Father: University	2.97	
Mother: No degree	21.71	
Mother: Primary	55.25	
Mother: Junior secondary	14.72	
Mother: High secondary or university	8.31	
Father: Blue coll, self-empl, not empl	74.74	
Father: White collar	17.4	
Father: Manager or entrepreneur	7.86	
Survey year: 1995	17.28	
1998	19.37	
2000	11.37	
2002	10.52	
2004	11.9	
2006	11.42	
2008	10.55	
2010	7.59	
N. of observations	6,604	

Note: binary variables are without standard deviation. Differently to other surveys, where the occupation of the father and the mother refers to the period in which the individual was teenager or in the childhood (like 15 years old), in the Bank of Italy Survey the same information is less precise as it refers to the status of the parent when he or she was the same age of the interviewed individual

MODEL AND ESTIMATION STRATEGY

As discussed in Section 2, our empirical model has one equation for education and one log wage equation with random coefficients. This is also called the correlated random coefficient model by Heckman and Vytlacil (2001). In the version proposed by e.g. Card (1999, 2001), the relationship between log wages and education S can be expressed as:

$$\ln W_i = \alpha_i + b_i S_i + u_i, \text{ where } \alpha_i = \alpha + X_i \beta + a_i.$$

When education is treated as continuous (years of schooling), this is the popular one-factor human capital model, where (individual) returns to education are constant across education levels and each additional year of education has the same marginal return. If education is treated as a discrete variable, the associated coefficient is the average of what one would expect to gain from going one step up in the schooling degrees ordered scale.

This way to formalise the structural equation of log-wages is a rather ad-hoc because the assumption that the individual return is independent of the level of attained education rules out the possibility that both observable and unobservable returns to schooling vary with the level of education. But this simplification has the virtue of tractability. One can of course specify a more flexible model, such as the generalised Roy with different potential wage equations for each education level (see Blundell et al., 2005). This implies that both observed and unobserved returns to education are different across schooling categories. Given that both average returns and comparative advantages vary with schooling, a model of this kind would account for a much more heterogeneity, but at the cost of many parameters to be estimated. We experimented with that, but in the end we decide to use a more tractable specification, which however captures the core of heterogeneity as it separates the schooling effect of pre-existing attributes (individual heterogeneity - ability) captured by a person specific intercept, from the unobservable gain from any additional year of education. Willis and Rosen (1979) develop a similar model for the binary treatment case with a dummy for college education under the normality assumption.

A general way to specify the average marginal return in the population (average treatment effect of an additional year of education) is the following: $b(X) = E(b_i | X) = b_0 + Xb_1$, where, for each additional year or level of education, the average return has a fixed component (intercept shift) and individual-specific returns to covariates. The latter captures that in principle part of the heterogeneity in returns can be due to observable characteristics (interacted terms). The whole range of individual returns can be written as follows: $b_i = b_0 + Xb_1 + e_i$ where, in addition to the common part and the one due to observables, we also have an idiosyncratic part due to comparative unobservable gains (or losses) from each unitary shift in education. We can therefore re-write the earnings equation as:

$$\begin{aligned} \ln W_i &= \alpha + X_i \beta + [b_0 + X_i b_1 + e_i] S_i + a_i + u_i \\ \ln W_i &= \alpha + X_i \beta + b_0 S_i + (X_i S_i) b_1 + (e_i S_i + a_i + u_i) \end{aligned}$$

Clearly, if we suspect that education is not exogenous, in principle we have a whole set of endogenous variables (basic variable plus interaction), and we would need at least as many instruments as variables, which is unfeasible in many practical applications. As a result, in our baseline specification with years of schooling we forget about the interactions and estimate the following model, with an homogeneous average marginal return and where heterogeneity is due to specific gains from additional education:

$$\ln W_i = \alpha + X_i\beta + b_0S_i + (e_iS_i + a_i + u_i) \quad [1]$$

The model with interactions will be estimated in our second specification, where we treat education as a discrete and ordered variable.

The two potential sources of bias in estimating the model by OLS are the ability bias and the return bias. The former is due to the correlation between a_i and S_i . The latter implies a correlation between the sorting gain e_i and S_i . And we cannot forget that as education is measured with some noise, there is another potential component in the error term, which is the measurement error. If e_i was the same or zero for all individuals, we are back to the standard homogeneous returns model with ability bias, which has motivated a large body of IV studies:

$$\ln W_i = \alpha + X_i\beta + b_0S_i + (a_i + u_i) \quad (2)$$

Heckman (1978) proposed a control function estimator based on normality assumptions to estimate a specific version of this model – the so-called dummy endogenous model –, where education is for example captured by a dummy for college degrees.

In general, it is reasonable to assume that both heterogeneity components exist as they proxy for influences by family background, preferences, ability. And it is also reasonable to think that people with higher benefits from education stay longer at school, meaning that the school variable is influenced by its own coefficient leading to self-selection due to sorting gains. If returns are heterogeneous based on unobserved correlated individual gains, IV are an unattractive alternative. As explained in Section 2, this motivates the use of an alternative approach based on the control functions (CF), to model the stochastic relationship between the two heterogeneity components a_i and e_i and S and Z . Z is a vector of variables correlated with S but which are mean-independent of unobserved ability and taste components.

To this purpose, let us specify a model where realised education levels (S) are function of an underlying reduced form latent process (S^*) for the optimal schooling level which guarantees the highest net utility level, expressed as the difference between benefits and costs from the schooling investment.

Desired education levels are influenced both by observable and unobservable factors like individual preferences, attitudes toward the risk, personal characteristics, family background. Actual education S is some transformation of S^* . In the case of years of schooling, the mapping is linear and maybe subject to measurement errors. The resulting reduced form is:

$$\begin{aligned} S_i^* &= Z_i\phi + \varepsilon_i \\ S_i &= t(S_i^*) = Z_i\gamma + v_i \end{aligned} \quad (3)$$

In the case of ordered levels of attained levels, we assume that. *EDU* is an ordered variable that takes *k* values from 0 to *J* depending on *EDU** crossing a set of given thresholds:

$$\begin{aligned} S_i^* &= \gamma Z_i + \varepsilon_i \\ S_i &= t(S_i^*) = k \text{ iff } \mu_k < EDU_i^* \leq \mu_{k+1} \\ k &\in \{0, 1, \dots, J\}, \quad \mu_k < \mu_{k+1} \quad \forall k \\ \mu_0 &= -\infty, \quad \mu_J = +\infty \end{aligned} \quad (4)$$

Z is the vector of schooling explanatory variables, which includes *X*. In the ordered probit case, the unobserved individual effect influencing human capital accumulation ε is distributed as a standard normal.

To correct the OLS bias we need an expression for the conditional mean of the error term (leaving implicit the dependence on *X*):

$$E(e_i S_i + a_i + u_i | S_i, Z_i) = E(e_i | S_i, Z_i) S_i + E(a_i | S_i, Z_i)$$

If we assume that the two conditional expectations are linear (as it is for example when wages and education are normally distributed), we can always specify the population regression of e_i and a_i as a function of the error term in the education equation. In the case of continuous educations we have:

$$\begin{aligned} E(e_i | S_i, Z_i) &= \theta_1 v_i, \quad \theta_1 = \text{cov}(e_i, v_i) / \text{var}(v_i) \\ E(a_i | S_i, Z_i) &= \theta_2 v_i, \quad \theta_2 = \text{cov}(a_i, v_i) / \text{var}(v_i) \end{aligned}$$

In the estimation, the error term is replaced by the first stage residual of the reduced form for education. The thetas are basically the covariances between the unobservables driving education and the sources of unobserved heterogeneity in wages. The wage equation to be estimated is:

$$\ln W_i = \alpha + X_i \beta + b_0 S_i + \theta_1 \bar{v}_i S_i + \theta_2 \bar{v}_i + u_i \quad (5)$$

In the specific case with homogeneous returns to education, the CF method is analogous to IV, since substituting first stage predictions of *S* in place of actual years as in the 2SLS procedure or adding the first stage residual gives exactly the same point estimates. When education is modelled by an ordered probit, the correction term involve the generalised order probit residual. In addition, it is possible to consistently estimate all the interactions between schooling and the other covariates without any additional instrument or complication (see Section 2).

The coefficients of the correction terms have a nice economic interpretation as they provide a direct test for the presence of 'selection bias' (endogeneity) and of 'return bias' induced by sorting gains (self-selection). If θ_2 is positive, selection in education is positive: given their high-

er ability, the more educated would have earned more even with the lowest schooling level. If it is negative, it may signal measurement error problems.

In principle, θ_1 may be positive or negative. Its sign helps to understand which theory – comparative wage advantages or compensating wage differences – better explains the behaviour of more or less educated individuals. As discussed above, individuals choose the schooling which gives the highest utility (monetary and non monetary factors). By assuming that unobserved factors are a proxy for motivation, good matches and other productivity-related wage determinants, if $\theta_1 > 0$ the comparative advantage of getting an additional year of education is higher for people who are intrinsically more likely to study more: by studying more, they receive higher expected monetary gains and are more productive. Since the allocation of workers is based on productivity-based comparative advantages, it is also efficient.

If $\theta_1 < 0$, the idiosyncratic component of the return is lower for people who are more likely to study more. Hence, studying is not motivated by monetary advantages. One possibility is that non monetary gains received more than compensate wage losses due to low unobserved productivity and comparative disadvantages. This would occur if, for example, education is perceived as a consumption more than as an investment goods. Or if higher qualitative job determinants raise satisfaction at the workplace and more than compensate unobserved (potential) negative wage differences. Since workers possess the education level in which they are not necessarily more productive, this allocation may be not considered efficient.

Specification and identification

The baseline specification of the wage equation is standard and comparable to what have been used in the previous studies for Italy. It includes basic controls for experience and its square, as well as dummies for the geographic area of residence, for the size of the municipality and for the survey year. Estimates with potential experience in place of actual experience deliver similar results but are less precise. Our final specification also fits the data better than the one with age and its square or with age only¹⁰.

We also estimate an extended specification, which is our preferred one and includes a list of variables for the schooling level of both parents and for the occupation of the father.

The reduced form for education includes all the wage regressors plus a number of excluded restrictions which are used to identify the model. We follow the previous literature for Italy and use the 1962 reform of the junior high school as a source of exogenous variation for education. Brandolini and Cipollone (2002) suggest that the individuals directly affected by the reform are those who were pupils between 6 and 14 years of age in 1963, i.e. born between 1949 and 1956. Among them, we can identify three groups, characterised by different exposures to the reform: the 1949 and 1950 cohorts were 13 and 14 in 1963. At least some of them were compelled to stay at schools longer than planned, but this displacement effect was probably small. The 1951 and 1952 cohorts were more directly exposed since they have just finished primary school at the

¹⁰ The final specification does not include dummies for the worker's occupation: the treatment of occupation dummies in the earnings equation is problematic because they are likely to be themselves outcomes of education decisions, and, therefore, part of the returns to education (Angrist and Pischke, 2009).

time of the reform. Younger individuals (in the 1953-1956 cohorts) attended primary school at the time of the reform, so that they were only potentially involved. We account for this heterogeneity by using three different instruments: a dummy for 1949-1950 cohorts, one for people born in 1951 and 1952 and one for people born from 1953 to 1956.

In addition, we also include a third cohort dummy for individuals born after 1956. This is motivated by the fact that, first, if we assume that the (conditional) distribution of tastes, opportunity costs and benefits associated with education does not change over time, these individuals, who completed their education in the 'reformed' school, can be included in the group of 'treated' individuals. Second, Cipollone and Brandolini (2002) show that the implementation of the reform was rather progressive and poor up to the early '70s: enrollment rates in low secondary school increased substantially, but in the sixties more than 15% of pupils still dropped out without getting the corresponding degree. Since we observe only the highest schooling degree and not enrollment rates, a dummy for post 1956 cohorts is expected to capture the long term effects of the 1963 reform.¹¹

Since we have more exclusion restrictions than endogenous variables, for the restricted specification we check the validity of our restrictions with over-identification tests.¹²

MAIN RESULTS

Results are in Table 2 and include standard OLS and CF findings for our different specifications of the wage and education equations. Similarly to previous studies for Italy, OLS and CF results in columns (1) to (3) refer to the baseline specification in which we do not control for family variables in the wage equation but include them in the reduced form for education. This model is equivalent to e.g. Brunello and Miniaci (1999). In columns (4) to (6) we add parents' education and father occupation to the set of wage covariates.

First of all, the F-test on excluded variables tells that our set of instruments is valid, in the sense that it is relevant in the reduced form for education and able to explain a substantial share of variation in education. Hence the first condition for a valid instrument (relevance) is satisfied.

Moving on to the analysis of wage results, column (1) results show that average marginal returns to education in the specification without family controls are about 5.3%, in lines with the

¹¹ We also replicated our analysis by using the approach followed by Brandolini and Cipollone (2002), that is by assigning the post 1956 cohorts into the control group, which also includes the pre 1949 cohorts. The results are very similar. We also estimated a second specification with an expanded set of instruments, including the whole set of interactions between cohort dummies and family background variables, to allow also the reform to have an heterogeneous effect on the level of schooling in terms of family characteristics. In principle, our set of cohort dummies – especially that for individuals born after 1956 – may pick up age effects or general education trends, such as its increase following the Italian economic boom of the 50s-60s. The inclusion of disaggregated controls for parents' education and occupation, the addition of interactions and the presence of a set of survey year's dummies should help capturing changes over time in the distribution of tastes, abilities and other unobserved attributes thus making pre and post-reform cohorts more comparable. In practice CF results using this larger set of Z were equivalent to the ones obtained with the basic set of cohort dummies. For this reason they are not presented and available upon request.

¹² The test is run by regressing the residuals of the selectivity corrected wage equations in (9) both on the vector X and on the vector of instruments. The resulting LM test has an asymptotic chi-square distribution with degrees of freedom equal to the number of over-identifying restrictions (see Main and Reilly, 1993).

other studies for Italy (see Section 3). The findings in column (2) from the model with homogeneous returns and random intercept only are equal of what we would get with a standard IV-2SLS procedure and can be directly compared to existing IV estimates for Italy. As it happens in most studies, also at the international level, OLS results underestimate the returns to an additional year of education, which are 7.5% according to CF. These are values close to Colussi (1997) and Cannari and D'Alessio (1995) and somehow lower than Flabbi (1996) and Brunello and Miniaci (1999), and smaller than estimated IV returns for many studies on United States, for example (see Card, 1999).

The coefficient of the control function is negative and significant, which is consistent with CF returns to education higher than OLS ones. This suggests that marginal returns for a random individual in the population are higher than actual observed returns. One possibility is that the correction term picks up the correlation between 'ability' and education, so that selection is negative: individuals with higher absolute wage advantages (say, ability) would be less likely to get high education. But this may simply reflect an OLS downward bias induced by large measurement errors in education, as it is likely to be the case, given that they are imputed.

Another possibility is that individual returns are indeed heterogeneous, and what we are picking up is just a local effect for the sub-population of individuals affected by the reform, for which the returns to additional schooling are higher than for the population as a whole. But this may not be necessarily true: since completion of junior high school was the expected outcome of the 1962 reform the 'local' effect of the reform may not be larger than the average if more years of compulsory schooling raised somehow education up to the junior high school degree to individuals without any additional payoff.

However, the Sargan tests of over-identifying restrictions for column (2) rejects the validity of the identification strategy, so that at least some of the variables used as instruments cannot be excluded from the wage equations at reasonable levels of statistical significance.

As discussed in Section 2, family variables are likely to be correlated with unobservable wage determinants and therefore should be added to the model as in columns (4) to (6). Results from column (4) show that, on the whole, the higher is the education of the mother and of the father, the higher are the wages. Since the mother typically spends more time with his son during his childhood, her education level may be a proxy for the cultural environment of the family and of its effects on the 'ability' of the son. Also the occupation of the father matters: as compared to having a father manager or entrepreneur, other occupations pay less. The inclusion of background variables reduces the size of returns to schooling to 4.8%, leaving the coefficients of other variables mostly unaffected. Hence, when controls for the parental background are omitted, the coefficients of the education dummies combine true returns to individual investments with returns to socioeconomic characteristics of the family of origin and maybe also the portion of the 'ability bias' captured by family background.

Once parents characteristics are added to the wage model, the Sargan test relative to the CF specification in column (5) strongly support the identification strategy and the choice of exclusion restriction. Results can be now trusted more, but are in linewith the findings of column (2): the marginal return to schooling is 7.3% and the correlation between the error term and

education is still negative. Since in column (5) we have taken out from the error term many variables which should be highly correlated with ability, our interpretation is that probably the ability bias is not the main responsible for the OLS relative to CF downward bias, which is probably due for the most part to measurement errors. The role played by heterogeneous returns is investigated by column (6), which includes an additional correction term to control for that. Its coefficient – which is proportional to the covariance between the distribution of individual returns and of education - is positive and statistically significant, though small in magnitude. It means that overall there exists positive self-selection and comparative advantages in education: individuals who are more likely to study more, benefits from higher unobservable wage gains from each additional year of schooling.

As a result, the net gain for a random individual decreases from 7.3 to 6.9%. By combining results from columns (5) and (6) we can conclude the shift in marginal returns to education from 4.8% to 7.3% is partly attributable to heterogeneity and sorting gains, which on average increases the observed returns but to a smaller extent: from 7.3 to 6.9%. The more plausible explanation for what is left – the jump from 4.8 to 6.9% - is a measurement error bias which attenuates OLS estimates and it is controlled for in the CF setting.

One way to investigate more on that is to re-estimate the model in terms of education levels instead of imputed years of schooling. The former should be much more reliable than the latter and subject to a substantially less amount of coding and recolling errors. If this is the case, the non interacted correction term would pick up absolute advantages driving endogenously, if any.

CONCLUSIONS

In this paper we used Italian data for head of household male full time workers in the age interval 25 to 60 to estimate marginal returns to education by a model which allow these returns to be heterogeneous in the population. The instruments for education are cohorts dummies that, similarly to previous studies for Italy, exploit the effects of a major change in compulsory schooling introduced in Italy in 1962. Differently to previous studies, the set of wage regressors also include family characteristics to control for unobserved individual traits inherited or learned from the parent, which are typically correlated with wages and education

According to our estimates, there is no evidence that more educated employees benefit of average from an absolute wage advantage. Using a simple control function approach, estimates reveal a negative correlation between unobserved wage determinants and years of schooling. Interpreting this result to make inference on the whole population of workers requires particular care. global effect. While we cannot a priori exclude that this is due to negative selection in education due to the less 'able' individuals earning higher wages and getting more education, our feeling is that this is due to large measurement errors of the education variable. This is also supported by the finding that self-selection in education is positive, so that more able individuals study more partly because they receive larger unobservable gains from additional years of schooling than the average. While, on the one hand, on average we have not enough evidence to conclude that education is able to attract the 'better' individuals

(higher ability), on the other hand self-selection works in the right direction. From a policy perspective, the allocation of talents can be considered efficient.

A direct control for heterogeneous returns and the inclusion of family variables among the wage regressors are the two major innovations of this paper over the previous evidence for Italy. This is just the first step for a better understanding of the role of heterogeneity in returns to education, maybe with more flexible and sophisticated models. This is left to future research.

REFERENCES [TO BE REVISED]

- Angrist J.D. and Pischke S. (2009) *Mostly harmless econometrics: an empiricist's companion*, Princeton University Press.
- Blundell R., Lorraine D. and Sianesi B. (2005) 'Evaluating the effect of education on earnings: models, methods and results from the National Child Development Study', *Journal of the Royal Statistical Society*, 168: 473-512.
- Brunello G. (2002) 'Absolute risk aversion and the returns to education', *Economics of Education Review*, 21: 635-640
- Brunello G., Comi S. and Lucifora C. (2000) 'The returns to education in Italy: a new look at the evidence', *IZA discussion paper n. 130*.
- Brunello G. and Dustmann C. (1997) 'Le retribuzioni nel settore pubblico e privato in Italia: un paragone basato su dati microeconomici', in: Dell'Aringa C. (ed.) *Rapporto Aran sulle retribuzioni 1996*. Franco Angeli, Roma.
- Brunello G. and Miniati R. (1999) 'The economic returns to schooling for Italian men: an evaluation based on instrumental variables', *Labour Economics*, 6: 509-19
- Cannari L. and D'Alessio G. (1995) 'Il rendimento dell'istruzione: alcuni problemi di stima', *Temi di discussione Banca d'Italia*, n. 253.
- Card D. (1999) 'The causal effect of education on earnings'. In Ashenfelter O, Card D (eds) *Handbook of Labor Economics*, vol 3B: 1801-1862.
- Card D. (2001) 'Estimating the return to schooling: progress on some persistent econometric problems', *Econometrica*, 69: 1127-1160.
- Chiburis R. and Lokshin M. (2007) 'Maximum likelihood and two-step estimation of an ordered-probit selection model', *The Stata Journal*, 7: 167-182.
- Cipollone P. and Brandolini A. (2002) 'Return to education in Italy: 1992-97', *paper presented at the XIV EALE conference – Paris*.
- Colussi A. (1997) 'Il tasso di rendimento dell'istruzione in Italia: un'analisi cross section', *Politica Economica*, XIII: 273-294.
- Flabbi L. (1997) 'Investire in istruzione: meglio per lui o per lei?', *working paper n. 08*, Università Milano-Bicocca, Dipartimento di Economia Politica.
- Lee L.F. (1978) 'Unionism and wage rates: a simultaneous equation model with qualitative and limited dependent variables', *International Economic Review*, 19: 415-433.
- Maddala G.S. (1983) *Limited Dependent and Qualitative Variables in Econometrics*, Cambridge, Cambridge University Press.
- Main B.G.M. and Reilly B. (1993) 'The employer-size wage gap: evidence for Britain', *Economica*, 60: 125-42.
- Vella F. and Gregory M. (1996) 'Selection bias and human capital investment: estimating the rates of return to education for young males', *Labour Economics*, 3: 197-219.

Table 2: OLS and CF estimates of the log wage equation: homogeneous and heterogeneous returns to education

	Specification without family variables						Specification with family variables					
	OLS		Control Function				OLS		Control Function			
			Homog. Returns		Heterog. Returns				Homog. Returns		Heterog. Returns	
	(1)		(2)		(3)		(4)		(5)		(6)	
Inhwage	Coef.	t	Coef.	t	Coef.	t	Coef.	t	Coef.	t	Coef.	t
Years of schooling	0.053	39.84	0.075	30.35	0.075	30.2	0.048	32.14	0.073	15.05	0.069	14.1
Experience	0.025	12.28	0.025	12.55	0.025	12.69	0.026	12.88	0.025	12.37	0.025	12.58
Experience ² (*100)	0.001	-7.89	0.001	-6.55	0.001	-6.69	0.001	-8.27	0.001	-6.47	0.001	-6.78
Area: North-East	-0.034	-2.89	-0.036	-3.04	-0.033	-2.83	-0.035	-2.99	-0.036	-3.04	-0.033	-2.83
Area: Centre	-0.057	-4.52	-0.060	-4.81	-0.059	-4.75	-0.055	-4.35	-0.061	-4.92	-0.060	-4.77
Area: South	-0.099	-7.87	-0.089	-7.05	-0.088	-6.98	-0.092	-7.24	-0.090	-7.06	-0.088	-6.97
Area: Islands	-0.068	-4.38	-0.057	-3.72	-0.060	-3.93	-0.061	-3.89	-0.058	-3.7	-0.060	-3.9
20,000<size<400,000	0.025	2.74	0.008	0.91	0.007	0.81	0.016	1.79	0.008	0.91	0.008	0.86
size>400,000	0.026	1.51	-0.004	-0.21	-0.005	-0.31	0.002	0.13	-0.006	-0.33	-0.008	-0.45
Father: Primary							0.036	2.11	0.012	0.7	0.016	0.91
Father: Junior high school							0.001	0.03	-0.041	-1.85	-0.032	-1.44
Father: Secondary high school							0.019	0.68	-0.045	-1.52	-0.032	-1.07
Father: University							0.079	1.93	-0.010	-0.23	0.007	0.16
Mother: Primary							0.012	0.76	-0.007	-0.41	-0.004	-0.22
Mother: Junior secondary							0.049	2.41	0.022	1.05	0.027	1.29
Mother: High second. or univer.							0.069	2.51	0.009	0.29	0.019	0.65
Father: White collar							0.054	4.05	0.021	1.4	0.024	1.66
Father: Manager or entrepreneur							0.112	5.46	0.072	3.24	0.076	3.44
v1			-0.030	-10.42	-0.048	-10.32			-0.028	-5.52	-0.043	-7.33
v1*years of schooling					0.002	4.59					0.002	4.56
Sargan Test			Chi2(12)=0.99 p-value=0.0025						Chi2(3)=0.123 p-value=0.867			
F-test on excluded instruments			F(4, 6575)=118,24 p-value=0.000						F(4, 6575)=118,24 p-value=0.000			

Notes: Pooled data for 1995, 1998, 2000, 2002, 2004, 2006, 2008, 2010. N. observations: 6,604. t-statistics are robust to heteroskedasticity and autocorrelation. Excluded categories: 1995, living in the north west, mother without education, father without education, father blue collar, self-employed or not employed.