Rivista internazionale di scienze sociali, 2020, n. 1, pp. 85-103

HETEROGENEOUS RETURNS TO EDUCATION IN ITALY

Eliana Baici, Paolo Ghinetti*

ABSTRACT

The estimation of the return from an additional education unit is subject to the ‘‘ability bias’’, responsible for the ‘‘endogeneity’’ of education to wages, and the ‘‘return bias’’, induced by self-selection into education levels based on individual-specific unobservable gains. We use a control functions estimator that, differently to standard instrumental variables, can account for both selection mechanisms. It adds two corrections terms to the wage equation, one for each selectivity source. We use Bank of Italy data for Italian men in the 25-60 age interval. Identification uses a major reform that in 1962 increased years of compulsory schooling, implemented through a set of cohort dummies as exclusion restrictions. We find no evidence of absolute unobservable wage advantages for the more educated. Still, they are positively self-selected: they study more because their marginal return to education is higher than the average, due to specific unobservable gains from additional schooling.

Keywords: Wages, Education returns, Control functions, Italy.
JEL Classification: J31, I26, C3.
ISSN: 0035-676X (print); 1827-7918 (digital)
DOI: 10.26350/000518_000043

1. INTRODUCTION

In the empirical literature on returns to education, a standard (sometimes implicit) assumption is that the wage gain from any additional unit of education (one year of schooling) is the same for all individuals (Griliches, 1977). Under this assumption, if observed education correlates with earnings through unobservables (‘‘ability’’ bias), instrumental variables (IV) deliver consistent estimates of the average marginal return to education (see Card, 1999, for a review).

In their seminal contribution, Willis and Rosen (1979) firstly emphasized that the same education investment may generate different benefits across individuals, and that this may be relevant for education choices. In principle, there would exist a whole distribution of returns, one for each individual.

We acknowledge comments from participants to the XXIX Aiel Annual Conference in Pisa and to a seminar at the University of Verona, and financial support through FAR 2017 funds at the Università del Piemonte Orientale.

* Eliana Baici, Paolo Ghinetti, Dipartimento di Studi per l'Economia e l'Impresa, Università del Piemonte Orientale. Email Baici: eliana.baici@uniupo.it. Email Ghinetti: paolo.ghinetti@uniupo.it

In the standard mincerian wage model, differences in ability (levels) are captured by individual-specific baseline wage (intercepts). Differences in returns, by individual-specific returns from one additional year of education (slopes). If returns are heterogeneous, the standard IV model identifies the average return to education only if, for any given individual, his or her specific return is uncorrelated with schooling choices, i.e. if people on average do not anticipate their future wage gains associated to education choices. This excludes the possibility of self-selection in education, which is a behavioral assumption broadly not consistent with economic theory (Heckman, 1997). Explicitly accounting for heterogeneous returns is then crucial for causal inference whenever people could forecast of have some knowledge of their individual return when deciding the optimal education level. If people self-select into education levels based on unobserved returns or ‘‘sorting gains’’, a second source of bias, a ‘‘return bias’’, arises in OLS estimates.

We estimate wage returns to education in Italy by allowing for both ability and individual-specific slopes potentially correlated with education. None existing paper for Italy addresses this heterogeneity issue, and only estimates for the ‘‘homogeneous’’ model are available. To this purpose, we use the control functions (CF) estimator proposed by Garen (1984) and Card (2001). The CF approach is similar to IV, but it can correct earnings equations for both endegenity and self-selection, to obtain estimates of returns to schooling free of the ‘‘ability’’ and the ‘‘return’’ bias.

We use data for Italian men in the 25-60 age interval from the seven 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 that in 1962 increased the number of years of compulsory schooling, and use a set of cohort dummies and their interactions with key family background characteristics as exclusion restrictions.

The paper is organised as follows. Section 2 reviews the relevant literature for Italy. Section 3 discusses the empirical model, estimation techniques and the identification strategy. Section 4 describes the main features of the data. Section 5 presents the results. Conclusions follow in Section 6.

2. LITERATURE FOR ITALY

Several papers analysed returns to education in Italy, for the most part 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 typically summarised by imputed years of schooling based on the highest attained degree. The main limitation of this measure is that it is rather inaccurate and potentially subject to substantial measurement errors: because 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 school in the 1942-48 period, aimed at capturing exogenous variations in education induced by the second world war, and find similar results.

Later works used schooling reforms as sources of exogenous variation in education. Flabbi (1997) analyses yearly earnings with SHIW data for 1991[1]. 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[2]. The affected students are identified by a cohort dummy for people born since 1951[3]. The analysis of Flabbi (*ibidem*) is limited to the cohorts of individuals born between 1946 and 1962. Otherwise, the 1951 cohort dummy may capture things other 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 occur on average at older ages for the more educated. Using a rich set of controls, males have higher IV returns to

[1] 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.

[2] Until 1963, only primary school was compulsory for children aged 6 to 11. Since October 1 1963, the leaving school age is 14, thus additing 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.

[3] 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. 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.

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 65%

The institutional features of the 1962 reform were used also by Cipollone and Brandolini (2002). 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 show 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 10%. 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 alterative explanation, is that the marginal returns of individuals affected by an exogenous increase in education 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.

3. EMPIRICAL FRAMEWORK

3.1. *Model*

The starting point of many studies is the mincerian log wage equation:

$$\ln W_i = a_i + bS_i \tag{1}$$

The random intercept captures that different individuals earn different baseline wages (once the effect of education is netted out), i.e. what they would earn without education (which is a counterfactual for many of them – those with some education). The standard assumption is that the individual intercept (labor market ability) can be decomposed into a constant, common to all individuals; a part which varies across individuals but according to some observable characteristics; and a part which is individual-specific and unobservable. Accordingly: $a_i = a + X_i\beta + u_i$, and we can write (1) in a more convenient regression format:

$$\ln W_i = a + X_i\beta + bS_i + u_i \tag{2}$$

When education is 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 gives the same marginal return. If this is (a reasonable linear approximation of) the population model, the main assumption is that the marginal return of education b is constant across individuals.

In this framework, if S is suspected to be endogenous, then $E(u \mid X, S)$ or $\mathrm{cov}(u, S \mid X)$ is not zero and OLS are not consistent. For example, unobservable characteristics that affect education (“ability”) may have also an impact on the wage levels, whatever the education level is. This correlation involves education and the part of wages independent of education, without any linkage with returns from the education investment. This is obvious because, by assumption, the return is the same for everyone. This assumption is of course very restrictive by standard economic theory and cost-benefit considerations but has the virtue that, in this context, conventional IV provide consistent estimates of b.

A more realistic behavioral model would perhaps tell more about the relationship between individual schooling decisions and associated wage effects. A simple way to incorporate a more detailed description of wages formation and education choices is to allow for heterogeneous education returns across individuals:

$$\ln W_i = a + X_i\beta + b_iS_i + u_i \tag{3}$$

In addition to a random intercept, this model has a random slope for education. Insofar, different individuals can experience different returns from the same education investment. In the literature it is known as the “correlated random coefficient model” by Heckman and Robb (1985) and Heckman and Vytlacil (1998, 2001) (see also Card, 1999, 2001; Blundell, Lorraine and Sianesi, 2005). There are many obvious reasons why the slope of education may differ across individuals, including tastes, individual attitudes, ability, etc.. In principle, we may decompose the individual return into three parts: $b_i = b + X_i\delta + e_i$, where, in addition to the common part and the one due to observables, there is an idiosyncratic component due to comparative unobservable gains (or losses) from each additional unit of education. For simplicity, hereafter we abstract from the part due to observable characteristics X[4]. The implicit assumption made by most IV studies is that e_i (or at least its conditional mean) is the same or zero for all individuals.

[4] The downside is that in this way we neglect that a portion of the individual return to education could be better explained by observable characteristics This is often made for conveni-

The final form of the log wage equation becomes:

$$\ln W_i = a + X_i\beta + bS_i + (e_iS_i + u_i) \tag{4}$$

where $e_i = b_i - b$[5].

The individual-specific part of the return to education captures the "net comparative wage gain" of the *i-th* individual in getting his actual education level, with respect to the population average. Of course, since education is a choice variable, its optimal level maximizes expected net benefits, that include unobserved returns e_iS_i, which go in the error term.

3.2. *Estimation issues*

On the estimation side, both the individual specific intercept a_i and the slope b_i are random variables that may correlate with the education variable S. If a_i varies across individuals with different observed education levels, by (4) education and unobserved earnings determinants u_i are correlated. In general, this correlation stems from the cost component of education: for example, when, conditional on the set of X controls, earnings are correlated with individual traits that reduce the cost of acquiring higher education. This endogeneity problem is unrelated to the value of b_i, (or, more precisely, of e_i) but it introduces a bias in the estimation of b: the average difference in wages between individuals, who differ for one year of education, is not a measure of the average wage gain that any individual would expect from one additional year of schooling. This is the well-known "ability bias" or "selection bias".

On the other hand, the dependence of the marginal return to education b_i on education levels implies a correlation between e_i and S. For example, unobserved components of the marginal return to an additional year of schooling may be high-

ence, since, with endogenous education, the interactions between S and X would be additional endogenous variables, requiring additional exclusion restrictions, which are not available in many empirical applications.

[5] This specification has the disadvantage that the individual return to each year of schooling is independent of the level of attained education, ruling out the possibility that not only observable, but also unobservable returns to schooling vary with the level of education. However, this simplification has the virtue of tractability. A more general approach would be the generalised Roy endogenous switching model, where each year of schooling or education level is associated with different potential and counterfactual wage functions. As a result, both observed and unobserved returns to education are different across years of education (see Blundell, Lorraine and Sianesi, 2005). Given that both average returns and comparative advantages vary with schooling, a model of this kind would incorporate a higher degree of individual heterogeneity, but at the cost of many additional parameters to be estimated. We experimented with that, but, in the end, we decided 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 the individual intercept, from the unobservable gain from any additional year of education.

er for the more educated, who therefore benefit from a higher comparative wage advantage from studying more. If this is the case, schooling is affected by its own return, and there is positive self-selection due to ‘‘sorting gains’’. If, on average, this gain from unobservables is not zero within each education level (as it is in the population as a whole), OLS estimates suffer from a ‘‘return bias’’. Finally, if education is noisily measured, there is another potential component in the error term, which is the measurement error.

If returns are heterogeneous based on unobserved correlated individuals gains, instrumental variables are able to identify only ‘‘local’’ effects (LATE), i.e. the average return for the sub-population of individuals affected by the instrument. This makes OLS and IV estimates not directly comparable, as they refer to different populations.

A simple alternative to IV is the control functions (CF) estimator, which is robust also to the ‘‘return bias’’. Instead of trying to eliminate the bias as IV do, the CF estimator controls for it. These control functions are the mean of the error of the wage equation conditional to the endogenous variables.

In practice, the CF methods gives a (linear) functional form to the stochastic relationship between the two heterogeneity components u_i and e_i and S, conditional on X and Z. Z includes the vector of exogenous variables X plus one or more exclusion restrictions (instruments) correlated with S but mean-independent of unobserved ability and taste components.

In particular, let us assume that the realised education level (S) guarantees the highest net utility level (benefits minus costs from the schooling investment). It depends on both observable and unobservable factors like individual preferences, attitudes toward the risk, personal characteristics, family background.

In the case of years of schooling – our continuous measure of education – the mapping is linear and maybe subject to measurement errors. Its reduced form is:

$$S_i = Z_i\gamma + v_i \tag{5}$$

To correct the OLS for ability and return biases we need an expression for the conditional mean of the error in eq. (4). Leaving implicit the dependence on X):

$$\mathrm{E}(e_iS_i + u_i \mid S_i, Z_i) = \mathrm{E}(e_i \mid S_i, Z_i)\, S_i + \mathrm{E}(u_i \mid S_i, Z_i) \tag{6}$$

If we assume that the two conditional expectations are linear, we can always specify the population regression of u_i and e_i as a function of the education equation's error term, as follows:

$$e_i = \theta_1 v_i + \eta_i,\ \mathrm{E}(e_i \mid S_i, Z_i) = \theta_1 v_i,\ \theta_1 = \mathrm{cov}(e_i, v_i) / \mathrm{var}(v_i); \tag{7a}$$

$$u_i = \theta_2 v_i + \varphi_i,\ \mathrm{E}(u_i \mid S_i, Z_i) = \theta_2 v_i,\ \theta_2 = \mathrm{cov}(u_i, v_i) / \mathrm{var}(v_i). \tag{7b}$$

Since the theta coefficients of the correction terms are proportional to the covariance between the unobservables driving the education process and the two components of the outcome's error term, the CF approach provides consistent estimates of the average return b also in case of self-selection.

The CF approach is implemented through a two-step procedure, in which the residuals from the first step reduced form for education are added to the wage equation in the second step to get an estimate of the above conditional means (Card, 2001; Blundell, Lorraine and Sianesi, 2005). The wage equation becomes:

$$\ln W_i = a + X_i\beta + bS_i + \theta_1 v_i S_i + \theta_2 v_i + \epsilon_i \tag{8}$$

where ϵ is a zero conditional mean error term and in place of the true v's we use its first step estimate. With homogeneous returns to education, the CF method is analogous to IV-2SLS, since substituting first stage predictions of S in place of actual years as in the IV-2SLS procedure or adding the first stage residual produces the same point estimates.

The coefficients of the correction terms have a nice economic interpretation. They test for the relevance of ''selection'' (endogeneity) and ''returns from unobservables'' (induced by sorting gains) mechanisms. If θ_2 is positive and statistically significant, selection in education would be positive: given their higher ability, the more educated would have earned more even with the lowest schooling level. If it is negative, ''negative selection'' may be a signal of poor measurement (biasing OLS toward zero) of the education variable.

In principle, θ_1 may be either positive or negative. Suppose that unobserved factors u_i are a proxy for motivation, good matches and other productivity-related wage determinants. Then, self-selection would be positive if people choose the level of schooling associated with the highest wage gain, such that unadjusted OLS estimates of b overestimate the true marginal return. Negative self-selection would arise if, instead, jobs requiring high education levels are on average associated with high non-monetary returns, which more than compensate for lower wages. If the individual-specific sorting return is positive (negative), workers are allocated efficiently in the education group in which their productivity is higher (lower).

In the empirical analysis, we first use a standard specification of Eq. (8) with 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 or age and its square in place of actual experience deliver similar results but are less precise.

Our preferred specification builds up the baseline and also includes controls for the schooling level of both parents and for the occupation of the father. As noted by Card (1999), while it is true that the transmission of the social and economic status partly occurs through the accumulation and development of human capital, and abilities developed before entering the labour market, family characteristics may directly influence individual wages even after controlling for the schooling level of the sons[6]. Due to the intergenerational correlation in abilities and the effect

[6] 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 indepen-

of parents' ability on their own education level, as a general empirical rule Card (*ibidem*) 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 and would produce more reliable IV or CF estimates.

3.3. *Identification*

Popular ''instruments'' for schooling decisions are ''natural experiments'' such as policy reforms that change years of compulsory schooling or modify the access to specific curricula. From the individuals' standpoint, they are typically exogenous and reduce the marginal cost of schooling for students in specific cohorts without any residual effect on unobservable earnings determinants. Using education differences between individuals who attended school 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 outcomes and endogenous variables is stable over time.

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 inter-cohort growth in educational attainment. Allowing for these systematic differences is then important when the identification strategy pools together different cohorts.

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. Cipollone and Brandolini (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 a different exposure 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 had just finished primary school at the 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

dent 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.

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. 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. 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 '60s more than 15% of pupils still dropped out without getting the corresponding degree. Because 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[7].

Since we have more exclusion restrictions than endogenous variables, for the restricted specification we check the validity of our restrictions with over-identification tests[8].

4. 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. Each year the survey covered approximately 8,000 households, corresponding to around 21,000 individuals and 14,000 labour income earners. 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 not sensitive to the choice of a specific year.

---

[7] We also replicated our analysis using the approach followed by Cipollone and Brandolini (2002), assigning the post 1956 cohorts to 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 the reform having a 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.

[8] 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).

The survey is cross-sectional, but a rotating panel component was added in 1987. Its size increased over the years and in recent waves represent half of the overall sample (about 4,000 observations in 2002). Because of that, the same individual may appear in different years. We eliminate ‘‘multiple-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 our sample follows that of most studies reviewed in Section 2. We focus on male household heads only (approximately 2,600 in 2002) to avoid issues of female labour force participation and household formation[9]. We further restrict the analysis to full-time employees who work in the non-agricultural sector (937 observations), who are in the age interval 25-60. For e.g. 2002, these selection criteria reduce the sample to about 900 units. Older individuals were excluded to avoid endogenous retirement issues, which are problematic especially for the more educated[10].

The final sample also excludes observations with missing values for family background variables. For 2002 this implies that we have 700 observations valid for the empirical analysis.

The SHIW provides a measure of annual earnings inclusive of extra-time compensations and fringe benefits, and net of taxes and social security contributions[11]. Additional information is on the average number of hours worked per week and on the number of months worked per year. Similarly to other studies, we construct an estimate of hourly net wages (inclusive of fringe benefits), 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). As a robustness check, we also consider monthly earnings and find results in line with those obtained with hourly wages. Nominal wages have been transformed into real ones (2010 €) using the consumer GDP deflator.

In the SHIW, the educational structure is summarised by a set of dummies for the highest completed schooling level. Similarly to many studies for Italy, we impute years of schooling by the standard 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).

Additional information used in the empirical analysis includes dummies for the

[9] We exclude retired people, unemployed, self-employed and students.

[10] 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 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.

[11] We also experimented by estimating equations for log wages without fringe benefits and non-monetary compensations, but the results were in practice the same as those reported in Section 6.

highest level of education of the mother and of the father, and for the occupation of the father, as well as labor experience, age, geographical area of residence, size of the municipality of residence and time dummies. A description of the sample and of the variables, and summary statistics are in Table 1.

TABLE 1 – *Variables' description and mean*

Variable	*Mean*	*Std. Dev.*
Real Hourly wage	10.286	6.341
Experience	23.881	9.578
Years of Schooling	10.911	3.829
Schooling 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 = univerity	11.19	
8 = postgrad.	0.56	
Residence area: North-West	26.51	
North-East	21.84	
Centre	20.91	
South	21.41	
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	
1949 or 1950	6.75	
1951 or 1952	6	
1953 to 1956	14.01	
post 1956	56.86	
Father education: no degree	17.28	
Primary	51.29	
Junior high school	19.32	

Secondary high school	9.15
University	2.97
Mother education: No degree	21.71
Primary	55.25
Junior secondary	14.72
High secondary or university	8.31
Father occupation: Blue coll, self-empl, not empl	74.74
White collar	17.4
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 childhood (e.g. 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.

The average of imputed years of education is about 11, which corresponds to a three years vocational degree in secondary schools. The average hourly wage is about 10.2, while labor market experience is 24 years. Only 12% of sample units have his father with high secondary school or university degree, and, for mothers, this happens only in the 8.3% of cases. The individuals who belong to the cohort exposed to the compulsory school reform are about 27% of our sample. The pre-reform control group amounts to 16.4%, while individuals born after 1956 and thus experiencing only the ‘‘reformed’’ school system are 57% of the sample.

5. RESULTS

Results are in Table 2. We present standard OLS and CF findings for two different specifications of the wage and education equations. In the first, we do not control for family variables in the wage equation; but include them in the reduced form for education to ''aid'' identification. OLS and CF results are in columns (1) to (3), respectively. In columns (4) to (6) we add parents' education and father occupation to the set of wage covariates.

Baseline OLS results in column (1) show that the average marginal return to education in the specification without family controls is about 5.3%, in lines with the previous evidence for Italy (see Section 2).

The CF findings in column (2) assume homogeneous returns (no individual slopes). They are equivalent to IV-2SLS estimates. First step diagnostics are in the lower part of the column. The F-test on excluded variables shows that our instruments are relevant and jointly statistically significant in the reduced form for education. Results suggest that OLS underestimate the average return to an additional year of education, which is 7.5% according to CF. The negative (and significant) coefficient (-0.030) of the correction term ($v1$) is consistent with a CF return higher than the OLS one. An estimated average schooling return of 7.5% by CF is comparable to existing IV-2SLS estimates for Italy, such as Colussi (1997) and Cannari and D'Alessio (1995), and somehow lower than in Flabbi (1997) and Brunello and Miniaci (1999). At its face value, a negative ''ability'' bias would suggest that the marginal return for a random individual is higher than the observed one. One possibility is that there exists ''negative selection'' into education: for some unobservable reason, the more educated are sorted in the group of individuals with lower (than the population average) absolute unobservable wages (say, ability). Alternatively, the negative correlation between education and unobservable earnings 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.

In any respect, the Sargan tests of over-identifying restrictions for column (2) rejects the validity of the identification strategy: some 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 correlate with unobservable wage determinants and therefore should be added to the wage equation as in columns (4) to (6). Results from column (4) show that 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 her off-springs during their childhood, her education level may proxy the cultural environment of the family and of its effects on the cognitive ability of the son. Also the occupation of the father matters: compared to being manager or entrepreneur, other occupations pay less. The inclusion of background variables reduces the size of OLS returns to schooling to 4.8%, leaving the coefficients of other variables mostly unaffected. This suggests that when we omit parental background controls, the coefficient of years of schooling combines true individual returns with returns to socioeconomic characteristics of the family of origin and maybe also the portion of the ''ability bias'' captured by family background.

TABLE 2 – *OLS and Control Function 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)	
lnhwage	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
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
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
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
Junior high school							0.001	0.03	-0.041	-1.85	-0.032	-1.44
Secondary high school							0.019	0.68	-0.045	-1.52	-0.032	-1.07
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
Junior secondary							0.049	2.41	0.022	1.05	0.027	1.29
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
Manager or entrepreneur							0.112	5.46	0.072	3.24	0.076	3.44
*v*1			-0.030	-10.42	-0.048	-10.32			-0.028	-5.52	-0.043	-7.33
*v*1*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			

Once we add parents' characteristics to the list of wage covariates, the Sargan test of the CF specification in column (5) strongly supports the identification strategy and the choice of exclusion restrictions. Estimates are now more robust than in Column (2), but the marginal return to schooling is similar and equal to 7.3%. Also the correction term is similar to that of column (2). To the extent which family background variables (before in the error term and now controlled for) are individual ability proxies, this suggest that the OLS downward bias may not depend to a somehow counterintuitive ''negative selection in ability'' argument, but to measurement errors in education.

An alternative explanation for the differences between OLS and CF is that returns are heterogeneous, and we are not taking into account that. In this circumstance, CF would pick up a local effect for the sub-population of people affected by the reform.

According to the 1962 reform, the compliers are those individuals who were ''forced'' to complete junior high school instead of dropping out without its completion or who otherwise would have dropped out just after the end of the primary school. Since their optimal schooling choice without the reform would have been to drop out early from the schooling system, their marginal net return from additional years of schooling would not have been very high. If this was the case, the estimated 7.5% would probably underestimate the average effect in the population.

A more general treatment of heterogeneous returns is in columns (3) and (6), where a specific correction term controls for it. 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. This suggests that there exists positive self-selection and comparative advantages in education: individuals who 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%. Ignoring that returns are heterogeneous and that this is relevant for education choices would overestimate the true return by 0.4% because it neglects that individuals who gain more from the education investment are self-selected in the group with higher education. The distortion is not dramatic, but still not negligible and statistically significant. It amounts to about 4% for an education investment of 10 years. The more plausible explanation for the jump from 4.8 (OLS) to 6.9% is a measurement error bias which attenuates OLS estimates and it is controlled for in the CF setting.

6. CONCLUSIONS

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 allowing these returns to be heterogeneous in the population. Identification uses a set of 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, wage equations include family characteristics to control for unobserved individual traits inherited or learned from the parent, which are typically correlated with wages and education either directly or indirectly through ability and other unobserved individual wage determinants.

Using a simple control function approach, we found that the correlation between unobserved determinants of absolute wages and years of schooling is negative. While we cannot exclude that ‘‘negative selection’’ in education is due to the less ‘‘able’’ individuals earning higher wages and getting more education or to ‘‘local’’ average treatment effects, our feeling is that this is due to large measurement errors in the education variable. About self-selection in education on unobserved individual returns, we find evidence of positive sorting: the fact that the more educated study more is partly because their marginal return to education is higher than the average, due to specific unobservable gains from additional years of schooling. Controlling for both ability and return biases, our findings suggest that in Italy one additional year of education grants on average a 6.9% higher wage.

On the policy side, we do have not enough evidence to draw about the ability of education to attract the ‘‘better’’ individuals in terms of their absolute wage potential. Nonetheless, self-selection mechanisms seem to work in the ‘‘right’’ direction: the highly educated are those whose returns from additional years of education are higher. From a policy perspective, this would generate an efficient allocation of talents.

The main limitation is that we assume a constant return from any additional year of education. In this respect, our analysis is just the first step towards a more thoughtful treatment of heterogeneous returns to education. A more general approach would assume different (and individual-specific) returns to each education level, which is also typically subject to smaller coding and recalling errors. This would enrich the behavioral content of the empirical model and reduce measurement concerns from using imputed years of education. The treatment of these issues is in our future research agenda.

REFERENCES

Angrist J.D., Pischke S. (2009). *Mostly harmless econometrics: an empiricist's companion*, Princeton University Press, Princeton

Bjorklund A., Moffit R. (1987). The Estimation of Wage Gaing and Welfare Gains in Self-Selection Models. *Review of Economics and Statistics*, **69**, 42-49

Blundell R., Lorraine D., 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., Lucifora C. (2000). *The returns to education in Italy: a new look at the evidence*. IZA Discussion Paper, n. 130

Brunello G., Dustmann C. (1997). *Le retribuzioni nel settore pubblico e privato in Italia: un paragone basato su dati microeconomici*. In C. Dell'Aringa (Ed.). *Rapporto Aran sulle retribuzioni 1996*. Franco Angeli, Milano

Brunello G., Miniaci R. (1999). The economic returns to schooling for Italian men: an evaluation based on instrumental variables. *Labour Economics*, **6**, 509-519

Cannari L., 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 O. Ashenfelter, D. Card (Eds). *Handbook of Labor Economics*, vol. 3B, pp. 1801-1862

Card D. (2001). Estimating the return to schooling: progress on some persistent econometric problems. *Econometrica*, **69**, 1127-1160

Cipollone P., 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

Heckman J., Navarro-Lozano S. (2004). Using Matching, Instrumental Variables, and Control Functions to Estimate Economic Choice Models. *The Review of Economics and Statistics*, **86**(1), 30-57

Heckman J., Robb R. (1985). *Alternative Methods for Estimating the Impact of Interventions*. In J. Heckman, B. Singer (Eds). *Longitudinal Analysis of Labor Market Data*. Cambridge University Press, New York, pp. 156-245

Heckman J., Urzua S., Vytlacil E.J. (2006). Understanding Instrumental Variables in Models with Essential Heterogeneity. *Review of Economics and Statistics*, **88**(3), 389-432

Heckman J., Vytlacil E. (1998). Instrumental Variables Methods for the Correlated Random Coefficient Model: Estimating the Average Rate of Return to Schooling When the Return is Correlated with Schooling. *Journal of Human Resources*, **33**(4), 974-987

Heckman J., Tobias J., Vytlacil E. (2001). Four Parameters of Interest in the Evaluation of Social Programs. *Southern Economic Journal*, **68**(2), 210-223

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 University Press, Cambridge

Main B.G.M., Reilly B. (1993). The employer-size wage gap: evidence for Britain. *Economica*, **60**, 125-142

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

Willis R., Rosen S. (1979). Education and Self-Selection. *Journal of Political Economy*, **87**(5), S7-36