LABOR
LABORATORIO RICCARDO REVELLI
CENTRE FOR EMPLOYMENT STUDIES

Working Paper no. **82**

Primary and secondary effects in educational attainment in Italy
Effetti primari e secondari nell'istruzione in Italia

Dalit Contini, Andrea Scagni, Annette Riehl
Università di Torino

Primary and secondary effects in educational attainment in Italy

Effetti primari e secondari nell'istruzione in Italia

Dalit Contini, Andrea Scagni, Annette Riehl
Università di Torino
dalit.contini@unito.it

Keywords: equality of opportunity, school choices, educational policies

1. Introduction

It is often held that educational expansion narrows social inequalities within nations by promoting a meritocratic basis for status attainment, yet substantial research indicates that the relative advantages of elite children over children with less privileged background have changed little in the last decades (Hannum and Buchmann, 2003; Shavit and Blossfeld, 1993; Breen and Jonsson, 2000); on average higher status children perform better in school and attain higher educational levels. In this light, equality of opportunity (EOP) in education is still a highly relevant issue in the international educational policy agenda.

Class differentials in educational attainment are related in the sociological literature to *primary* and *secondary* effects (Boudon, 1974). The former refer to the influence of social origin on ability early in children's educational careers: high status parents are more likely to sustain and motivate the school work and provide a stimulating environment to their offspring. The latter operate through the choices that families make within the educational system (including exit) *given* the level of ability. The rational action approach (Goldthorpe, 1996; Breen and Goldthorpe, 1997), assuming that families wish to avoid intergenerational downward mobility, provides a theoretical explanation for the evidence that, at given levels of ability, school choices vary across social background. Ability is intended here as an observed measure of school performance (typically grade point average) as opposed to unobserved measures of cognitive abilities, since it is held that it is the former that affects the decision process through the perceived probability of schooling success.

The evaluation of primary and secondary effects is particularly relevant at the end of compulsory schooling, where in many countries students face the decision whether to enrol into the academic track[1] (which gives access to tertiary education), to enrol into a vocational track, or to enter the labour market.

EOP obviously depends on institutional features and can be affected by educational policies. Interventions aimed at enhancing the performance of children of less advantaged background are likely to help containing primary effects. Secondary effects can be reduced by endorsing the enrolment of lower status children into the academic track or, possibly, by regulating access through ability assessments.

The evaluation of the relative importance of primary and secondary effects is the aim

---

[1] The term *tracks* is often used in the literature to indicate the different secondary school educational paths available to students in a certain educational system. The *academic* track is the one conceived to prepare for university studies (even if in some countries it is not strictly necessary to enter tertiary education).

of a growing body of literature (Erikson *et al,* 2005; Jackson *et al.,* 2007; Stocké, 2007; Kloosterman *et al.,* 2007). This research – based on surveys carried out at a national level – provides empirical evidence of the relevance of secondary effects in the creation of class differential in educational attainment. The methodology, briefly sketched in Section 2, combines the estimates of the distribution of school performance and of the probability of choosing a specific track given school performance, at each level of social background. A counterfactual argument is carried out: for each j and $h,$ the probability of entering the academic track that individuals would face if they had the ability distribution of class j, but the transition probability given ability of class h, is evaluated. Observed and counterfactual odds-ratio are compared, and a decomposition based on counterfactuals provides an estimate of the relative importance of primary and secondary effects.

Aim of this paper is to provide an assessment of primary and secondary effects in secondary school choices in Italy. Other countries studies (UK, Sweden, Germany, Netherlands) are based on panel surveys recording data on children's schooling careers, but prospective longitudinal data is not available for Italy. The analysis is based on the data of the survey *Percorsi di studio e di lavoro dei diplomati* (ISTAT, 2004), which collects detailed information of individuals educational histories up to three years after the secondary school degree. A major issue to deal with is self-selection (see Section 4), as only secondary school graduates are interviewed[2]. By integrating the survey data with administrative and census information we derive estimates of the relevant distributions, correcting for selection bias.

As lower secondary school final marks are assigned on a 4-level scale (*satisfactory, good, very good, excellent*[3]), a semi-parametric version of the standard approach is adopted. Results are described in Section 5. The main conclusion is that secondary effects are more important in shaping social origin differentials in secondary schools decisions than primary effects. By comparing our estimates with those reported in the recent literature, it turns out that the relative contribution of primary effects is substantially weaker in Italy than in the other countries.

2. The methodology

Let A be a continuous measure of students' school performance before track choice and S a discrete variable representing social status. Then $f(A|S)$ is the distribution of the performance scores for each social group; assuming a normal distribution, the relevant parameters can be estimated by group sample mean and variance.

Define Y as a binary variable taking value 1 if the academic track is chosen and 0 otherwise (i.e. if the student chooses a different track or if he does not enter secondary education). Note that Y refers to the first choice after the end of compulsory schooling and not to possible subsequent changes. The transition probability *given* performance $P(Y = 1 | A, S)$ can be estimated with binary logistic regression for each class separately.

---

[2] Employing data from PISA (*Programme for International Student Assessment*; OECD, 2005) in this context would greatly weaken the sample selection problem, since students are interviewed at 15, i.e. near the beginning of secondary school. However this option proves impossible since PISA does not include information on students' performance before track choice. PISA may however be appropriate to evaluate the *total effect* of social background on track choice (see for example Contini and Scagni, 2008).

[3] As translated from *Sufficiente, Buono, Distinto, Ottimo* in Italian. Obviously, students can also fail the exam.

The integrals:

$$P_{jj} = \int_{-\infty}^{+\infty} f(A|S=j) P(Y=1|A, S=j) dA \tag{1}$$

evaluated for each S by numerical integration, represent the predicted probability $P(Y=1|S=j)$ whose observed counterpart is the percentage of those belonging to social class j enrolling into academic schools.

On the other hand, the integral:

$$P_{jk} = \int_{-\infty}^{+\infty} f(A|S=j) P(Y=1|A, S=k) dA \tag{2}$$

is a "counterfactual" probability. Expression (2) is the probability that an individual would experience if he had the performance distribution of social class j and the transition probability of class k. With K social classes, there are $K(K-1)$ counterfactual probabilities.

The total effect of class j over class k on the propensity to continue to the academic track is represented by the odds ratio:

$$Q_{jj.kk} = \frac{P_{jj}/(1-P_{jj})}{P_{kk}/(1-P_{kk})} \tag{3}$$

Define also:

$$Q_{jj.kj} = \frac{P_{jj}/(1-P_{jj})}{P_{kj}/(1-P_{kj})}$$

The numerator represents the odds of continuing to academic education for an individual exposed to the performance distribution and the transition probability of class j, while the denominator represents the odds for an individual with performance distribution of class k and transition probability of class j. Since the difference here lies only in the performance distributions, this quantity is informative on primary effects. Similarly:

$$Q_{kj.kk} = \frac{P_{kj}/(1-P_{kj})}{P_{kk}/(1-P_{kk})}$$

provides information on secondary effects, as what varies here is the transition probability while the performance distribution remains fixed.

The total effect (3) can be factorized in two distinct ways:

$$Q_{jj.kk} = Q_{jj.kj} \, Q_{kj.kk}$$

$$Q_{jj.kk} = Q_{jk.kk} \, Q_{jj.jk}$$

By taking the logarithms, we obtain:

$$L_{jj.kk} = L_{jj.kj} + L_{kj.kk} \tag{4a}$$

$$L_{jj.kk} = L_{jk.kk} + L_{jj.jk} \qquad (4b)$$

where in each case the first term on the right hand side refers to situations with different performance distribution but the same transition probability, and the second term to situations with the same performance distributions and different transition probability. The relative importance of secondary effects can be evaluated by $L_{kj.kk} / L_{jj.kk}$ or $L_{jj.jk} / L_{jj.kk}$. Estimates based on (4a) and (4b) generally differ, although in practice not to a great extent (see Erikson *et al*, 2005 for details).

Assuming that there are only two social levels to ease the understanding: *H* (high) and *L* (low), we obtain the following expressions:

$$L_{HH.LL} = L_{HH.LH} + L_{LH.LL}$$

where the log total effect is given by the primary effect evaluated with the transition probability of the *high* class and the secondary effect with the performance distribution of the *low* class;

$$L_{HH.LL} = L_{HL.LL} + L_{HH.HL}$$

where the first term is the primary effect evaluated with the transition probability of the *low* class and the second term is the secondary effect with the performance distribution of the *high* class.

It is worthwhile to note that under the linear probability model: $P(Y=1|A,S) = \mu + \lambda S + \theta A$, with $A = \alpha + \beta S + \varepsilon$, the following would hold:

$$P(Y=1|S=j+1) - P(Y=1|S=j) = \beta\theta + \lambda$$

In this case primary effects are represented by $\beta\theta$ and secondary effects by λ. Instead, it can be shown that under the logistic model (even in the absence of interaction effects between *A* ad *S*):

$$\ln \frac{P(Y=1|A,S)}{1-P(Y=1|A,S)} = \mu + \lambda S + \theta A$$

primary and secondary effects – as measured by means of (4a) and (4b) – are functions of the parameters of both the model for *A* and the model for *S*, although the component related to primary effects is much more sensitive to β and θ, and the component related to secondary effects is much more sensitive to λ.

3. The analysis for Italy

3.1 Institutional features

Although compulsory education starts at age 6 and ends at age 16, the last two years are a very recent formal requirement; for the cohort considered in this work (students born in 1982) the end was still set at 14. There are five years of primary school and three years of comprehensive lower secondary education, after which students choose their upper secondary school among many different programmes.

As in most European countries, in spite of the wide range of different secondary school types, a broad distinction between an academic and a technical/professional track can be made. The academic track lasts five years and includes different types of *lyceum*: *classical, scientific, linguistic, artistic*[4]. The technical and vocational tracks (lasting respectively five and three years) lead directly to a professional qualification.

There are no special admission requirements, such as ability tests or marks, to enter the different tracks. After five years of schooling (with two integrative years for vocational schools), all tracks give access to university (Eurydice, 2006b). In practice, only few students from the vocational track enter tertiary education: in the ISTAT sample little more than 20% did, while the proportion for lyceums was higher than 90% and about 50% for technical schools.

3.2 The data

Differently from other countries, no adequate panel survey recording schooling careers is available for Italy. Given this limitation, we employ cross-sectional data from the survey *Percorsi di studio e di lavoro dei diplomati* carried out by ISTAT on higher secondary school *graduates*, recording the relevant longitudinal information retrospectively. The survey takes place every three years since 1998, and graduates are interviewed three years after the degree attainment, with the aim to investigate the transition from secondary school to tertiary education or work. As we will point out in Section 4, the nature of the survey implies the existence of significant sample selection, which will have to be dealt with.

The survey on 2001 graduates (ISTAT, 2004) has been employed in this paper. The data has been collected with a two stage sampling scheme, on 1868 schools and 20408 individuals[5]. Essential to our analysis is the recording of the final student's mark at the end of lower secondary school, the subsequent track choice, and a set of variables describing parental occupational and educational status[6].

3.3 Final marks in lower secondary school

According to the rational choice theory (Breen and Goldthorpe, 1997), families make their educational choices with the aim to avoid downward mobility, according to future employment prospects and the probability of schooling success of their children relative to each option. This assessment is made by taking into account children's ability, conceived as an observed measure of school performance as opposed to unobserved measures of cognitive abilities.

In Italy, the final lower secondary school mark[7] is the main observed information on

---

[4] At the time of track choice for the ISTAT sample, the *Istituto magistrale* prepared for the primary school teaching career. Although this type of school has been later redefined as *socio-pedagogic lyceum,* and further university education is now required to enter the teaching profession, until a few years ago this school gave direct access to it. For this reason we will not consider this school type as belonging to the academic track.

[5] Interviews were carried out with CATI. Initial sample size was set to 29.000 units; telephone numbers were wrong or there was no reply in about 25% of the cases, and 4% ended up with refusals.

[6] Although we do not employ this classification in the present paper, the data allow to define individuals with respect to three social classes as in the simplified British *National Statistics Socio-Economic Classification,* used for example in Jackson *et al.* (2007).

[7] The mark is attributed after a national exam, detached from normal school activity, held at the and of

children's ability before track choice. We highlight three possible sources of measurement error:

(i)   Final lower secondary grades encompass in Italy only four distinct proficiency levels (*excellent, very good, good, satisfactory*). This highly discrete grading system appears to be quite a rough measurement of students' ability when compared with other countries marks, based on finer scales (e.g. ten levels in the British case).

(ii) Exams are set up by the school teachers, and are not based on standardised national tests[8]. An indirect evidence of the existence of a bias is that, although international assessments such as PISA (OECD, 2005) show a significantly lower average level in Southern Italy with respect to the North, in the South the percentage of *excellent* is higher than in the rest of the country.

(iii) Related to point (ii), if marks were given with some reference to the average ability within the school, higher performing schools could evaluate their students somewhat more severely. The issue is particularly relevant in highly socially segregated environments, since on average high status children perform better.

The problem of measurement error is not explicitly addressed here. The reason is twofold. First, we think that the main source of bias in the Italian case is likely to be given by sample selection, due to employing data on secondary school graduates.

Perhaps more importantly, the second reason has to do with the rationale of the analysis. If it is true that people make their educational choices on the basis of observed school performance[9], the "correct" measure of ability for secondary effects is given by marks, even if they are affected by measurement error. On the other hand, the "correct" measure for identifying class differentials in the performance distribution should be latent ability.

Nevertheless, it is important to note that the decomposition method described above involves investigating the role of manifest ability in *shaping school choices*. In fact, social class transition probabilities - see formula (1) - are a weighted average of the class transition probabilities given ability (marks), where weights are given by the relative proportion of individuals with each level of ability (again, marks) within the class. In this light it is not relevant whether the school mark is a measurement error version of true ability (measurement error is instead very relevant for the assessment of inequality of opportunity in the true ability distribution across social classes). Thus, when we come to interpret primary effects in this context, we should acknowledge that what is here called "primary effects" has to do with the distribution of latent ability *and* the way this ability is actually translated in marks[10].

Yet, this caveats would not hold if people were aware of their true level of ability and shaped their decisions accordingly: transitions rates would have to be assessed given true ability and weights defined consequently. If marks were employed in this case, the relative contribution of primary effects would be underestimated[11].

---

lower secondary school (*Esami di Stato conclusivi del I ciclo*).

[8] This issue is likely to become less relevant in the future: from 2007, in fact, final exams include two standardized tests (linguistics and mathematics) with common evaluation guidelines.

[9] Stocké (2007) addresses this issue for Germany and finds that educational choices are driven mainly by school marks, although a minor effect can be ascribed to parents' perception of their children ability.

[10] It is nevertheless obvious that in the extreme case where marks were hardly related to ability, the decomposition itself would loose much if its meaning, in that secondary effects would become the only source of class differentials.

[11] With the aim to investigate this issue we have developed a simulation study (not presented here). The

4. Sample selection

As we have pointed out, no adequate panel survey recording school careers is available for Italy, and for this reason we employ the ISTAT cross-sectional survey on graduates 2001, recording the relevant information retrospectively. Since the survey target population does not include those who have enrolled into a secondary school and exited the educational system before graduation[12], the sample is affected by selection bias.

We now deal with the consequences of sample selection on the estimates of the relevant distributions for primary and secondary effects. We will show that without corrections, we would *underestimate* both the differences in the ability distribution across social background levels, and the effect of social background on school choices. Note that traditional methods for dealing with sample selection (e.g. Heckman's method) cannot be employed in this context because micro-data on dropouts is not available.

Primary effects

As before let A be the school performance before track choice and S a measure of families social status. Define G as a binary variable taking value 1 if the child has attained a secondary school degree and 0 if he has dropped-out of the educational system. The distribution of interest is $P(A|S)$, while the observable distribution is $P(A|S, G=1)$. The two distributions are related by:

$$P(A|S,G=1)=P(A|S)\frac{P(G=1|A,S)}{P(G=1|S)}=P(A|S)\frac{P(G=1|A,S)}{\int_A P(G=1|S,A)P(A|S)dA}$$

The observable distribution and the distribution of interest coincide if the second factor in the right hand side is equal to 1, i.e. if performance A does not affect the graduation probability given social status. Since this is obviously very unlikely, the survey estimate of the performance distribution given social status is biased.

Let us recall that in the Italian system ability is measured on a 4-level ordinal scale, which we will code as: *satisfactory* (1), *good* (2), *very good* (3), *excellent* (4). We will make the assumption that school drop-outs come exclusively from the population of low performers (see next section for empirical evidence on this):

$$P(G=0|S,A=j)=\begin{cases} f(S)>0 & if\ j=1 \\ 0 & if\ j=2,3,4 \end{cases} \tag{5}$$

For $j=2,..,4$ this implies that:

$$P(A=j|S,G=0)=\frac{P(G=0|A=j,S)P(A=j|S)}{P(G=0|S)}=0 \tag{6}$$

Since:

$$P(A|S)=P(A|S,G=1)P(G=1|S)+P(A|S,G=0)P(G=0|S) \tag{7}$$

---

by combining (6) and (7) we obtain:

$$P(A=j|S)=\begin{cases} P(A=j|S,G=1)P(G=1|S) & if\ j=2,3,4 \\ 1-\sum_{j=2}^{4}P(A=j|S) & if\ j=1 \end{cases} \tag{8}$$

In order to estimate $P(A|S)$, we employ the ISTAT graduates' survey to assess $P(A|S, G=1)$, but we also need to estimate the graduation probability given social status $P(G=1|S)$. Since:

$$P(G=1|S)=\frac{P(S|G=1)P(G=1)}{P(S)} \tag{9}$$

we will estimate $P(S|G=1)$ from the graduates survey and exploit the official statistics derived from administrative data sources for the overall graduation probability $P(G)$ and the social status distribution $P(S)$ (see Section 5.2).

Secondary effects

Let Y represent again secondary school choice: $Y=1$ for the academic track and 0 otherwise. We are interested in $P(Y=1|A,S)$, but we can only estimate $P(Y=1|A,S,G=1)$. Since:

$$P(Y=1|A,S,G=1)=P(Y=1|A,S)\frac{P(G=1|Y=1,A,S)}{P(G=1|A,S)} \tag{10}$$

the survey estimate is unbiased if, given ability and social status, the graduation probability does not depend on the chosen track. Note that Y refers to the *first* choice undertaken at the end of compulsory school, while graduation can be achieved in *any* track. Students may change track if they fail or if they are not satisfied with their initial choice, and then graduate. In this light, the likelihood to attain a secondary school degree will not depend on how difficult or selective the specific track is. The enrolment into the academic track will be considered instead as a signal of higher *aspirations*.

The consequences of employing directly the graduates' survey to estimate $P(Y=1|A,S)$ can be easily grasped by assuming the simple linear probability models:

$$P(G=1|A,S,Y)=\alpha+\beta A+\gamma S+\delta Y$$

$$P(Y=1|A,S)=\lambda+\xi A+\theta S$$

We obtain:

$$P(G=1|A,S)=P(G=1|A,S,Y=1)P(Y=1|A,S)+P(G=1|A,S,Y=0)P(Y=0|A,S)$$

$$=(\alpha+\beta A+\gamma S+\delta)P(Y=1|A,S)+(\alpha+\beta A+\gamma S)P(Y=0|A,S)$$

$$=\alpha+\beta A+\gamma S+\delta P(Y=1|A,S)$$

Then the second factor in the right hand side of (10) is:

$$\frac{P(G=1\,|\,Y=1,A,S)}{P(G=1\,|\,A,S)} = \frac{(\alpha+\beta A+\gamma S)+\delta}{(\alpha+\beta A+\gamma S)+\delta(\lambda+\xi A+\theta S)}$$

This expression is never smaller than 1 (it is equal to 1 if δ=0), and is a decreasing function of A and S. In fact, $(\lambda+\xi A+\theta S)<1$ (since it is a probability); given that parameters are positive, it is an increasing function of A and S. Thus, the observed probability is greater than the probability of interest for all status, but it is increased by a greater factor for the lower social background. As a consequence, secondary effects are *underestimated*. As we will show in the next section, by employing a different data source, empirical evidence suggests however that δ should be nearly 0, meaning that aspirations are entirely captured by school performance and social status. No corrections are needed in this case; $P(Y=1|A,S)$ is estimated directly from the graduate's survey data.

4.1 Supporting the assumptions

We now turn to data analyses carried out in order to provide empirical support to the assumptions described in the previous section.

Primary effects

Let us recall the relevant assumption, described by equation (5), stating that, for each social background, only low performers eventually drop-out from school. The marginal distribution of performance can be written as:

$$P(A = j) = P(A = j\,|\,G = 1)P(G = 1) + P(A = j\,|\,G = 0)P(G = 0)$$

from which we obtain the performance distribution for school-drop-outs:

$$P(A = j\,|\,G = 0) = \frac{P(A = j) - P(A = j\,|\,G = 1)P(G = 1)}{P(G = 0)} \tag{11}$$

This distribution can be roughly estimated by employing the graduates survey data – providing information on $P(A\,|\,G = 1)$ – and aggregate administrative data from ISTAT – which records the overall distribution of lower secondary final examination marks $P(A)$ for the year 1996, as well as an estimate of the overall national percentage of school dropouts $P(G=0)$ for the same year. Applying (11) we obtain:

$\hat{P}(A = 1\,|\,G = 0) = 0,96 \qquad \hat{P}(A = 2\,|\,G = 0) = 0,05$

$\hat{P}(A = 3\,|\,G = 0) = 0,005 \quad \hat{P}(A = 4\,|\,G = 0) = -0,02$ [13].

strongly supporting the assumption.

Secondary effects

We now evaluate the assumption:

---

[13] Small inconsistencies among the combined data sources produce a negative probability, which is however so close to 0 to be reasonably considered negligible.

$$\frac{P(G=1\mid Y=1,A,S)}{P(G=1\mid A,S)}=1 \tag{12}$$

i.e., that the effect of aspirations is entirely captured by that of school performance and social status.

As we have pointed out before, no longitudinal micro-data on schooling careers is available for the estimation of the conditional distribution of *G*. However, a survey carried out jointly by CISEM and IARD[14] in 2006 on 3600 upper secondary school students in the area of Milan can be employed for this purpose. The sample includes students in each of the five grades of the upper secondary schools; information on school careers as well as family characteristics, including parental educational and occupational status are recorded. The survey is cross-sectional and does not include dropouts; nevertheless, by comparing 1° grade students (including all future dropouts) with 5° grade students (assuming that nobody exits the school system thereafter), we can roughly assess the dropouts profile.

By a simple application of Bayes' theorem[15]:

$$\frac{P(G=1\mid Y=1,A,S)}{P(G=1\mid A,S)}=\frac{P(Y=1\mid G=1,A,S)}{P(Y=1\mid A,S)}$$

The right hand-side can be estimated by the ratio of the proportion of the academic track students in 5° grade and that for 1° graders. Considering *S* as the highest parental education and modelling both $P(Y=1\mid G=1,A,S)$ and $P(Y=1\mid A,S)$ with binary logit regressions, these ratios are all close to 1[16], supporting the validity of (12).

5. The empirical analysis

5.1 Semi-parametric approach

In Section 2 school performance *A* was taken as a continuous variable – as in most countries marks follow a fine scale and in some cases the grade point average is employed – which can be approximated quite well by a normal distribution; although not strictly necessary, this is also useful for the numerical evaluation of integral (1).

Because of the highly discrete scale (see Section 3.3), the normal distribution is clearly not appropriate for Italy. In this context, let *A* be the discrete variabile taking values 1 to 4, corresponding to the four proficiency levels from lowest to highest. Expression (1) becomes:

$$P_{jj}=\sum_{A=1}^{4}P(A\mid S=j)\,P(Y=1\mid A,S=j) \tag{13}$$

---

[14] CISEM stands for *Centro per l'Innovazione e Sperimentazione Educativa Milano* and is a research centre on educational problems of *Provincia di Milano*. IARD - Istituto Franco Brambilla is a research centre focusing on life problems and opportunities of young people. The authors would like to thank both CISEM and IARD for the collaboration and availability of data.

[15] Since $P(G=1\mid Y=1,A,S)\;=\;\dfrac{P(G=1,Y=1\mid A,S)}{P(Y=1\mid A,S)}\;=\;\dfrac{P(Y=1\mid G=1,A,S)\,P(G=1\mid A,S)}{P(Y=1\mid A,S)}$.

[16] These ratios vary from 0.93 for high status-high ability students to 1.32 for low-status-low ability students.

and counterfactual probability (2):

$$P_{jk} = \sum_{A=1}^{4} P(A|S=j) P(Y=1|A,S=k) \qquad (14)$$

The performance distribution *P*(*A*|*S*) is estimated non-parametrically, given gender and geographical area (*North West*, *North East*, *Center*, *South and Isles*); although the transition probability $P(Y=1|A,S)$ could be estimated non-parametrically as well, to privilege parsimony and keep results simple we employ binary logit models as in the original approach.

Note that although in the relevant literature social class - derived from parental occupation (Erikson and Goldthorpe, 1992) - is generally considered, for the moment we operationalize *S* with reference to the highest parental educational attainment. The main reason is that, having to correct for sample selection by employing national aggregate statistics on official reports, there seems to be stronger coherence between the two data sources[17]. In what follows, the terms indicating social origins will always refer to this concept.

5.2 Sample selection correction factors

As we have seen in Section 4, in order to correct for sample selection, for the evaluation of $P(A|S)$ we need to estimate $P(G=1|S)=P(S|G=1)P(G=1)/P(S)$. The three factors in the right hand side have been obtained separately for males and females and for each geographical area as follows:

- $P(S|G=1)$ has been estimated directly from the graduates survey data;
- $P(G=1)$ is the marginal graduation probability for the cohort under study; it has been computed as the ratio of the number of graduates in 2001 (data directly obtained from the Education Ministry Statistical Office) to the number of births in 1982 (data from ISTAT *Annuario di Statistiche Demografiche* official publications).
- *P*(*S*) is the national distribution of the highest parental educational level for the 1982 birth cohort (the 19 years old in 2001), derived from ISTAT *Annuario di Statistiche Demografiche* (which reports the joint parental educational level at child birth).

The first estimates of $P(G=1|S)$ were not always strictly smaller than one and mild inconsistencies in their ordering were found[18]. This could be due to the fact that we employ different data sources, which are likely to be affected by non-sampling error in various ways[19]. Another potential source of bias is that the data employed for the estimation of *P*(*S*) refers to parental educational level at child birth, and does not record the educational qualifications attained afterwards; to be consistent with the definition of status in the survey the distribution at age 14 should be employed.

Since these first estimates varied little among geographical areas and inconsistencies were found to be weaker on aggregate data, the distributions were evaluated at a

---

[17] Note also that for Italy the odds ratio between *Y* and *S* when status is measured by social class is much lower than that relative to the highest parental educational level. Moreover, some recent works seem to be going in the same direction (see e.g. Kloosterman *et al.,* 2007).

[18] Students from upper level families appeared in few cases to be slightly more likely to to drop out than those from middle level families.

[19] Sampling variability should enter here only via $P(S|G=1)$, but the standard errors of the estimates from the graduates survey are very small, and cannot alone explain the inconsistencies.

national level. Furthermore, to take into account the slight bias towards lower educational levels due to employing parental education at child birth, small adjustments to the estimates of $P(G=1|S)$ were applied[20].

The final estimates are reported in Table 1 below.

Table 1. *Estimated probabilities $P(G=1|S)$ of attaining the upper secondary school diploma by parental educational level and gender*

Parental education	tertiary	upper secondary	lower sec./primary
males	0.87	0.85	0.52
females	1.00	0.99	0.60

5.3 Results

Following the approach outlined in Section 4.1, primary and secondary contributions to $P(Y=1|S=j)$ are evaluated. A sketch of the procedure is outlined in Figure 1.
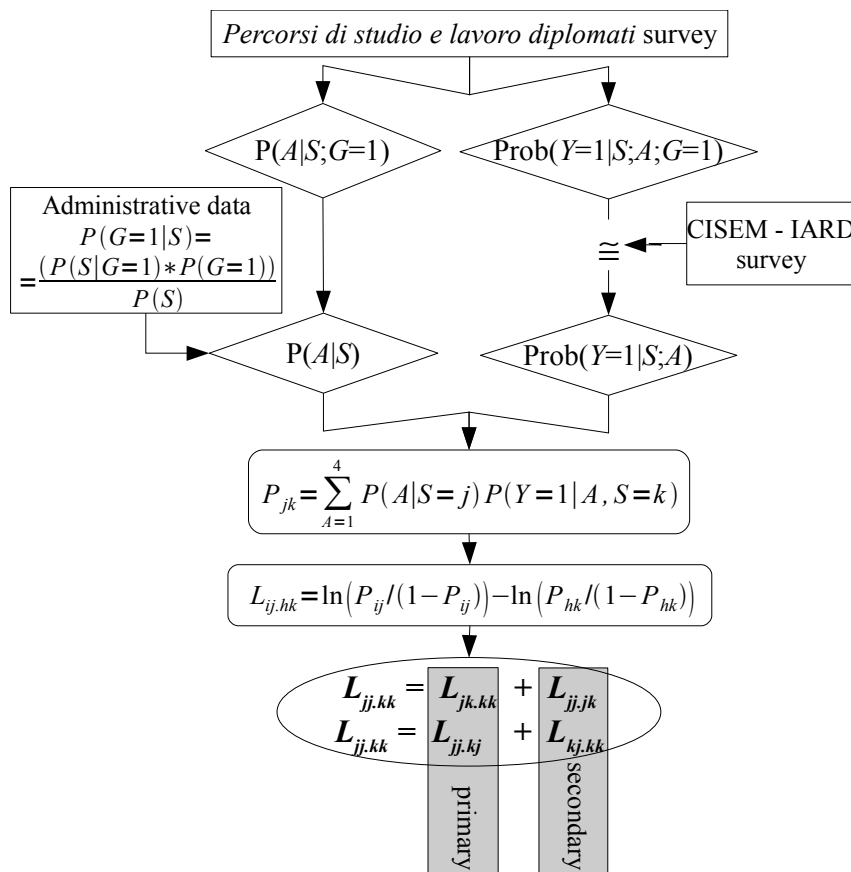
Figure 1. *Scheme of the applied decomposition technique*

---

[20] We assume somewhat arbitrarily that 5% of those with upper secondary qualification obtain a university degree after their child birth and that 7% of those with compulsory education obtain a high school diploma. With these reasonably sized values all probability estimates did not exceed 1. Note that decompositions (4a) and (4b) are only slightly affected by mild changes in these percentages.

$P(A|S)$ – conditional on highest parental educational level, but also on gender and geographical area – is estimated from the graduates survey and corrected for sample selection as in (8) and (9). Results are reported in Table 2. The distribution is much more favourable for children from well educated families. Moreover we observe that females are better performers than males, in line with the evidence from all over the world, and that more positive marks are observed in the South and Isles[21].

Table 2. *Lower secondary school final mark distribution P(A|S) after sample selection correction, by highest parental educational level, gender and area*

		Male					Female			
	Parental education	Lower sec. school final mark				Parental education	Lower sec. school final mark			
		satisfactory	good	very good	excellent		satisfactory	good	very good	excellent
North-West	tertiary	0.26	0.23	0.25	0.26	tertiary	0.12	0.23	0.22	0.43
	upper sec.	0.38	0.29	0.17	0.16	upper sec.	0.17	0.29	0.29	0.25
	lower sec./prim.	0.68	0.17	0.10	0.05	lower sec./prim.	0.56	0.23	0.13	0.08
North-East	tertiary	0.27	0.30	0.25	0.18	tertiary	0.06	0.25	0.29	0.40
	upper sec.	0.42	0.28	0.19	0.11	upper sec.	0.17	0.33	0.27	0.23
	lower sec./prim.	0.71	0.16	0.08	0.05	lower sec./prim.	0.58	0.22	0.12	0.08
Centre	tertiary	0.28	0.26	0.18	0.28	tertiary	0.10	0.17	0.28	0.45
	upper sec.	0.40	0.28	0.18	0.14	upper sec.	0.21	0.27	0.24	0.28
	lower sec./prim.	0.72	0.16	0.06	0.06	lower sec./prim.	0.58	0.21	0.12	0.09
South Isles	tertiary	0.25	0.17	0.23	0.35	tertiary	0.06	0.17	0.23	0.54
	upper sec.	0.39	0.26	0.18	0.17	upper sec.	0.18	0.23	0.22	0.37
	lower sec./prim.	0.68	0.18	0.08	0.06	lower sec./prim.	0.53	0.20	0.13	0.14

Table 3. *Raw transition rates to the academic track P(Y=1|A,S) by highest parental educational level, lower secondary school final marks, gender and area*

		Male					Female			
	Parental education	Lower sec. school final mark				Parental education	Lower sec. school final mark			
		satisfactory	good	very good	excellent		satisfactory	good	very good	excellent
North-West	tertiary	0.37	0.61	0.80	0.90	tertiary	0.29	0.61	0.82	0.84
	upper sec.	0.13	0.30	0.47	0.77	upper sec.	0.07	0.25	0.50	0.67
	lower sec./prim.	0.04	0.10	0.14	0.46	lower sec./prim.	0.03	0.12	0.28	0.48
North-East	tertiary	0.31	0.58	0.78	0.96	tertiary	0.21	0.61	0.72	0.91
	upper sec.	0.08	0.20	0.43	0.73	upper sec.	0.08	0.24	0.43	0.64
	lower sec./prim.	0.03	0.07	0.18	0.59	lower sec./prim.	0.03	0.10	0.21	0.56
Centre	tertiary	0.51	0.68	0.84	0.92	tertiary	0.32	0.55	0.83	0.93
	upper sec.	0.10	0.26	0.51	0.72	upper sec.	0.15	0.26	0.53	0.74
	lower sec./prim.	0.03	0.09	0.20	0.45	lower sec./prim.	0.06	0.21	0.34	0.50
South Isles	tertiary	0.36	0.68	0.69	0.86	tertiary	0.23	0.42	0.68	0.85
	upper sec.	0.10	0.21	0.39	0.63	upper sec.	0.13	0.22	0.50	0.68
	lower sec./prim.	0.02	0.08	0.28	0.47	lower sec./prim.	0.09	0.13	0.28	0.51

---

[21] As noted in Section 3, there is evidence of some measurement error across the country, as this result is not in line with the standardised resulted by the international assessment PISA.

Table 3 shows the raw observed transition rates to the academic track for all sub-groups. As expected, the propensity to enrol into a *liceo* is much higher among better performing students and for those from higher status.

As anticipated in Section 4.1, in order to compute counterfactual probabilities and the ensuing decomposition into primary and secondary effects, $P(Y=1|A,S)$ was modelled with binary logit regression on A, gender and area - separately for each value of S. Preliminary log-linear analysis showed no significant interactions among these regressors, so only main effects of the three covariates were included in the model. Since the parameter estimates for the distinct A levels are remarkably close to a progression with unit steps for all three educational levels, models where A was taken as a quantitative covariate (taking values 1-4 from *satisfactory* to *excellent*) were preferred. The results are shown in Table 4.

At the same level of demonstrated ability the transition probability is much higher among high status children[22]. The effect of ability is approximately the same among the different parental background. Gender differences are less marked; the variable is significant for lower and upper S: females with low educated parents are more likely than males to enter the academic track given ability, while transition probabilities are higher for males from families with tertiary education. Geographical effects are not very clear.

Table 4. *Logit models for the transition probabilities to academic track*

	S = tertiary			*S* = upper secondary			*S* = lower sec./primary		
Full model	β	Sig.	Exp(β)	β	Sig.	Exp(β)	β	Sig.	Exp(β)
ind_North-West	0.286	0.014	1.33	0.208	0.001	1.23	-0.147	0.101	0.86
ind_North-East	0.252	0.053	1.29	-0.034	0.646	0.97	-0.196	0.055	0.82
ind_Center	0.594	0	1.81	0.256	0.000	1.29	0.126	0.162	1.14
gender (female)	-0.290	0	0.75	0.069	0.168	1.07	0.385	0	1.47
ind_buono	1.030	0	2.8	0.955	0	2.6	1.050	0	2.86
ind_distinto	1.827	0	6.22	1.976	0	7.22	2.032	0	7.63
ind_ottimo	2.765	0	15.88	2.895	0	18.09	3.088	0	21.94
constant	-0.778	0		-2.242	0		-3.314	0	
Simplified model	β	Sig.	Exp(β)	β	Sig.	Exp(β)	β	Sig.	Exp(β)
ind_North-West	0.289	0.013	1.34	0.210	0	1.23	-0.149	0.097	0.86
ind_North-East	0.259	0.046	1.3	-0.031	0.669	0.97	-0.197	0.053	0.82
ind_Center	0.595	0	1.81	0.257	0	1.29	0.125	0.164	1.13
gender (female)	-0.287	0	0.75	0.069	0.169	1.07	0.385	0	1.47
A	0.899	0	2.46	0.971	0	2.64	1.023	0	2.78
constant	-1.614	0		-3.213	0		-4.327	0	
No geographic area	β	Sig.	Exp(β)	β	Sig.	Exp(β)	β	Sig.	Exp(β)
gender (female)	-0.288	0	0.75	0.072	0.15	1.07	0.382	0	1.47
A	0.871	0	2.39	0.965	0	2.62	1.026	0	2.79
constant	-1.315	0		-3.098	0		-4.365	0	

---

[22] This can be seen from the raw probabilities in Table 3 and is reflected in the values of the constant in the logit models in Table 4.

Factual and counterfactual probabilities P_{ii} and P_{ij} estimated as in (13) and (14) are reported in Table 5[23]. Rows refer to school mark distributions according to parental education, while columns indicate which level of S is used to model the transition probability to the academic track. Thus, the numbers located on the diagonal are the actual estimated transition probabilities $P_{jj} = P(Y=1|S=j)$ for each family status. These values are higher for females; stronger gender differences are observed for low S: females with low parental education are almost twice as likely to enrol into a *liceo* than males.

Off diagonal elements P_{jk} are instead counterfactuals, combining lower secondary school marks distribution and conditional transition probabilities for different parental educational levels. For example, $P_{11}=0.677$ is the transition rate of a male whose parents have tertiary education. The transition probability of an hypothetical child with the ability distribution of the upper class but the conditional propensity to choose *liceo* of the lower class is given by $P_{13}= 0.206$; similarly, the transition probability when the ability distribution is that of the lowest class and the conditional propensity is that of the upper class is $P_{31}= 0.489$.

Table 5. *Estimates of P_{ij} for Italy*

	Male P(Y=1|S;A) referring to...			Female P(Y=1|S;A) referring to...		
P(A|S) referring to...	tertiary	upper sec.	lower sec. /primary	tertiary	upper sec.	lower sec. /primary
tertiary	0.677	0.382	0.206	0.726	0.506	0.353
upper secondary	0.597	0.290	0.142	0.644	0.411	0.272
lower sec. /primary	0.489	0.190	0.082	0.480	0.251	0.149

There is a noticeable tendency - somewhat stronger for males - to decline faster along rows than along columns, indicating that the differences in family preferences for $Y=1$ due to S *given* children's marks are more relevant in determining the track choice with respect to school performance differences due to S.

Table 6 presents the results of the decomposition into primary and secondary effects. Both formulas (4a) and (4b) are computed, and produce similar results; average contributions are also reported. The main finding is that *secondary effects* tend to prevail in all contexts, the sole exception being that of medium vs. low status females.

It is important to recognise that this result does not imply that class differentials in children's *ability* are weak (see the discussion on measurement error in Section 3), nor that differentials due to S in children's *school marks* are weak. Results imply instead that *differentials due to S in secondary school choices* are mainly driven by differences in the transition probabilities given previous school performance, while differences in the performance distributions play a weaker role. This may occur either because performance distributions vary little across social status, or because performance does not affect much school choices[24]. Distinguishing between these two alternatives is

---

[23] Since the estimates resulting from the simplified models without the dummies for area do not change much with respect to the ones coming from the extended model, the more parsimonious specification was employed

[24] In principle, there could be wide family status differences in the observed level of ability, but if school choices are affected little by performance, depending mainly on social status, these differences would not exert a relevant role.

possible by looking directly at the estimates of $P(A|S)$ and $P(Y=1|A,S)$.

By comparing the estimates between males and females and across social origins, we can see that the relative importance of secondary effects is stronger for males than for females, and is stronger when comparing upper and middle status with respect to middle and low status. With respect to gender, by looking at Table 2 we find no clear differences in the social status effect on the performance distribution[25]. Furthermore, from Table 4[26] we derive that the social origin effect on the probability to choose the academic track given ability is milder for females than for males. Given that the effect of ability is very similar across values of *S*, we may conclude that the gender difference in the *relative* contributions of primary and secondary effects is due to weaker secondary effects for girls (in *absolute* terms) rather than to stronger primary effects.

Table 6. *Primary and secondary effects decomposition*

	Male			**Female**		
	tertiary > upper sec.	tertiary > lower sec./primary	upper sec. > lower sec./primary	tertiary > upper sec.	tertiary > lower sec./primary	upper sec. > lower sec./primary
$L_{jj.kk}$	1.634	3.156	1.522	1.333	2.720	1.387
$L_{jk.kk}$	0.347	0.782	0.556	0.381	1.054	0.736
$L_{jj.jk}$	1.287	2.374	0.966	0.952	1.666	0.652
% primary	0.212	0.248	0.365	0.286	0.387	0.530
% secondary	0.788	0.752	0.635	0.714	0.613	0.470
$L_{jj.kj}$	0.416	1.068	0.621	0.384	1.140	0.761
$L_{kj.kk}$	1.218	2.089	0.901	0.949	1.581	0.626
% primary	0.255	0.338	0.408	0.288	0.419	0.549
% secondary	0.745	0.662	0.592	0.712	0.581	0.451
average % primary	0.234	0.293	0.387	0.287	0.403	0.539
average % secondary	0.766	0.707	0.613	0.713	0.597	0.461

6. Conclusions

The results described in Section 5.3 are particularly interesting when considered within the international context. The most striking finding is that the relative contribution of *primary effects* is *much lower in Italy* than in the other countries for which the analysis has been carried out. Let us review the main results. Primary effects[27] account for about 76% of the total social background effect in UK (Jackson *et al.*, 2007, for year 2001), 58% in Stockholm, Sweden (Erikson, 2007, for 1990), 47% in the German Lander Rhineland (Stocké, 2007, for 2003), 58% in the Netherlands (Koosterman *et al.*, 2007, for 1999). The corresponding estimates for Italy are much lower: 29.3% for males and

---

[25] Moreover, by estimating, somewhat improperly, a linear model for performance, we do not find significant interaction effects between gender and status, (i.e. the effect of status on performance does not change with gender).

[26] See the constant and the gender coefficient.

[27] The percentage with respect to the high-low status comparison is reported here.

40.3% for females. Although these values are not fully comparable, because of cross-country institutional differences, definitions of social status[28] and because ability assessments are not always standardized, differences are however large, and it would be of great interest to understand the reasons laying behind them.

We can think of different topics for further work:

(i) In order to interpret the results from a comparative point of view, the *absolute* contributions of primary and secondary effects should be evaluated together with the *relative* ones. This implies recovering comparable estimates of the total effect of social origins on school choices. Note however that cross-country comparisons are even more problematic in this case: employing parental education or social class can give rise to substantials differences within countries[29].

(ii) The low importance of primary effects in Italy with respect to other countries can have two alternative interpretations: a) social background differentials in the school performance distributions are relatively weak; b) the role of ability in educational decisions is weak. Given the difficulties in cross-country comparisons based on national data, evidence from the international assessment carried out on 4th graders, PIRLS (*Progress in International Reading and Literacy Study;* Mullis *et al.*, 2003) can help to shed some light on this issue. Simple regression analysis indicate for example that Italy is one of the countries with the *lower* inequality of opportunity with respect to performance scores near the end of primary school.

(iii) The assessment of how specific institutional features – in particular, early tracking – affect equality of opportunity in education is the focus of an interesting body of work (Hanushek and Woessman, 2006; Woessman, 2007; Brunello and Checchi, 2007): by employing international surveys like PISA, the school design effect is identified by exploiting the cross-country variability. To our knowledge no attempt has been done yet to deepen the understanding of how institutional features promote or discourage primary and secondary effects[30]. In order to put forward educational policies with the aim to reduce educational inequality it would be very useful to try opening the *black box* and separate the effects on school performance from those on choices given performance. At the moment this aim is difficult to accomplish, as on one hand it is difficult to harmonise national data to allow for adequate cross-national comparisons, on the other hand international data such as PISA cannot be employed for this purpose, because no measure of ability before school choice is available. This could be an interesting challenge for future research.

---

[28] In UK and Sweden father's social class, in Germany mother's social class, in the Netherlands and Italy the highest parental educational level.

[29] We can see this from PISA, for which common alternative definitions are possible. Taking the highest parental educational level the following raw OR between high and low social status are found: Netherlands 4.7, Italy 6.9, Germany 12.9. Taking social class, Netherlands 8.5, Italy 5.8, Germany 8.4.

[30] For example, why is it that in Italy primary effects are so low? Could it be due to the fact that the compulsory school system is quite highly standardised in Italy? (standardization refers to the degree to which the quality of education meets the same standards nationwide; Allmendinger, 1989). On the other hand, secondary effects are strong. Is this related to the absence of performed-based restrictions to the academic track, at work in other countries (Netherlands for example)?

References

Allmendinger J. (1989) *Career Mobility Dynamics: A Comparative Analysis of the United States, Norway, and West Germany*, Berlin: Max-Planck-Institute fur Bildungsforschung.

Boudon R. (1974) *Education, Opportunity and Social Inequality.* New York, Wiley

Breen R., Goldthorpe J.H. (1997) Explaining educational differentials. Towards a formal rational action theory, *Rationality and Society*, Vol 9(3), 275-305

Breen R., Jonsson J.O. (2000) A Multinomial Transition Model for Analyzing Educational Careers, *American Sociological Review* 65:754-772.

Brunello G., Checchi D. (2007) Does school tracking affect equality of opportunity? New international evidence, *Economic Policy*, 52

Checchi D., Flabbi L. (2007) Intergenerational mobility and schooling decisions in Germany and Italy: the impact of secondary school tracks, *IZA Discussion Paper* 2876

Contini D., Scagni A. (2008) Equality of Opportunity in Secondary School Enrollment. Comparing Italy, Germany and the Netherlands, unpublished manuscript

Erikson R. (2007) Social Selection in Stockholm Schools: Primary and Secondary Effects on the Transition to Upper Secondary Education, pp. 61-81 in: *From Origin to Destination. Trends and Mechanisms in Social Stratification Research*, edited by Stefani Scherer, Reinhard Pollak, Gunnar Otte, and Markus Gangl, Campus Verlag, Frankfurt

Erikson R. and Goldthorpe J.H. (1992) *The constant flux: a study of class mobility in industrial societies*, Oxford, Clarendon Press

Erikson R., Goldthorpe J.H., Jackson M., Yaish M., Cox D.R. (2005) On class differentials in educational attainment, *Proceedings of the National Academy of Sciences,* vol. 102, n. 27, 9730–9733

Eurydice, Italian Unit (2006b) *Eurybase the Information Database on Education Systems in Europe,The Education System in Italy 2005/2006*, http://www.eurydice.org

Goldthorpe J.H. (1996) Class analysis and the reorientation of class thoery: the case of persisting differentials in educational attainment, *British Journal of Sociology,* 45(3): 481-506

Hanushek E.A., Woessman L. (2005) Does educational tracking affect performance and inequality? Difference-in-differences evidence across countries, *IZA Discussion Paper n.* 1901

Hannum E., Buchman C. (2003) The consequences od global educational expansion, *Occasional paper of the American Academy of Arts and Sciences*, Cambridge, MA

ISTAT (2004) *Percorsi di studio e di lavoro dei diplomati*, Roma

Jackson M., Erikson R., Goldthorpe J.H., Yaish M. (2007) Primary and Secondary Effects in Class Differentials in Educational Attainment: the Transition to A-Level Courses in England and Wales, *Acta Sociologica,* 50(3), 211-229

Klosterman R., de Graaf P., Ruiter S., Kraaykamp G. (2007) Parental education and the transition to higher secondary education. A comparison of primary and secondary effects for five Dutch cohorts (1965-1999), *Proceedings of the RC28 Spring Meeting in Brno*, Czech Republic

Mullis I.V.S., Martin, M.O., Gonzalez, E.J., & Kennedy, A.M. (2003) *PIRLS 2001 International Report: IEA's Study of reading literacy achievement in primary schools*, Boston College

Shavit Y., Blossfeld, H.P. (1993) (eds.) *Persistent inequality: changing Educational attainment in thirteen countries*, Westview, Boulder, Colorado

Stocke V. (2007) Strength, Sources, and Temporal Development of Primary Effects of Families´ Social Status on Secondary School Choice, Sonderforschungsbereich 504, WP series, n° 07-60

Woessmann L. (2007) Fundamental Determinants of School Efficiency and Equity: German States as a Microcosm for OECD Countries, *CESifo Working Paper* 1981