Migrant students classroom allocation policy in Italian schools

Andrea Scagni – University of Torino

Abstract When evaluating an educational system, its ability to seamlessly integrate children of different ethnic backgrounds is today an important issue. We propose a method to test in what part of Italian primary schools children are allocated to classes without paying attention to their ethnic background.

1 Introduction

The recent development of international assessments of student learning has led to a fertile research activity on performance determinants. Some studies focus on socio-demographic factors, others on the impact of school features. Comparative analyses usually focus on the design of school systems. The evaluation of the *net school effect* has attracted a strong interest, aiming to "measure" the ability of the specific school to improve the education of its students. This would help to allocate resources based on efficiency criteria, helping to focus research of the factors that contribute to differentiate "good" and "bad" schools.

Among the factors linked to net school effects there are *peer effects*, i.e. the influence on individual student performance of the classroom environment, where all educational activities take place (and which is also the main social interaction context together with the family). Most studies agree that the socio-economic composition of the class affects performance, while only limited work deals with the composition in terms of immigrant background (Contini, 2011). A seemingly well established result is that class composition effects are stronger for less advantaged children.

The student body composition in a school cannot be easily affected by public policy, but schools are mostly free to allocate new students among classes as they deem appropriate (no official guidelines on this exist in Italy). In this process they could take into account some children characteristics, such as migrant status, parental job status and education, and judgments of the teachers of previously attended schools. In the extreme, classes could be formed grouping children that are strongly homogeneous with respect to such characteristics or, on the opposite, with no regard for the latter. This second option could be seen as technically equivalent to a *random allocation* of children within classes.

In recent years the incidence of migrant students in classrooms has rapidly emerged as a hotly debated issue, as the number of migrant children enrolled in Italian primary schools increased significantly, and their uneven spatial allocation (mostly in poor urban areas) has in some extreme cases implied that native students are now a minority. Most opinions agree that, at least up to a certain degree of concentration, it is best for the whole community to avoid the "segregation" of migrant students in dedicated classes, as peer effects help integrating them both in terms of school performance as well as of general social interaction. Lately, however, opposite opinions have also gained some ground.

In this work, based on INVALSI 2009/10 SNV data, the allocation of children with *migrant status* among schools is assumed as given, and we try to assess if, within each school, they tend to be scattered equally among all classes, resulting, when numbers are not too small, in a proportional allocation mechanism, or are instead allocated following some different scheme, e.g. clustered in only a few classes. The preliminary analysis carried out in this paper involves the 5^{th} graders in primary school.

2 The method

Let k_j be the number of classes in school j (j = 1, 2, …., S, where S is the number of schools of a certain level in a given country), and n_{ij}, i = 1, 2, …., k_j the size of each class, which is here assumed to be given. Further, let m_{ij} be the number of migrant students allocated in class i of school j, so that n_{ij} - m_{ij} is the number of native students in it. It is assumed that the total number of migrant students enrolled in the school $\sum_{i=1}^{k} m_{ij}=M_j$ and the total size of school j $\sum_{i=1}^{k} n_{ij}=N_j$ are given.

In this context we try to test if class composition in Italy is consistent with a random allocation model *at the system level*. Second, if this hypothesis is rejected, we assume that there is a part of the school system for which it holds, while another part uses other criteria to allocate students in classes. In this case, we would like to assess the *size* and *composition* of the two parts and, possibly, if schools in the two sets differ in some ways.

A simple approach to the problem treats the test on each school as a chi-squared test of independence X^2_j, comparing observed frequencies of a k_j x 2 table (m_{ij} ; $n_{ij} - m_{ij}$) against their expected values under H_0 $\left(\frac{M_j \cdot n_{ij}}{N_j}; \frac{(N_j - M_j) \cdot n_{ij}}{N_j}\right) \quad \forall i, j$. Since a sum of independent χ^2 variables has the same distribution, $\sum_{j=1}^{S} X^2_j$ is asymptotically a χ^2 with $\sum_{j=1}^{S} k_j$ d.f. under H_0. Some of the problems involved in the analysis concern low frequencies - quite common since many schools deal with only a handful of migrant students, H_1 being completely unspecified without any role in the analysis, and the fact that this is a sum of local test rather than an actual global test.

Due to this, robustness of results is checked applying also a generalized $LR = L_0 / L^*$ (likelihood ratio) test, where $L_0=\prod_{j=1}^{S} L_{j0}$ is the likelihood under H_0 of observing a certain allocation of migrant children in all schools, assuming independence among schools. When $k_j = 2$, in the 2 x 2 table where marginals assumed to be fixed on both side (class sizes and migrant/native totals), the exact distribution under H_0 is of the hypergeometric type. For $k_j > 2$ this generalizes to the *multivariate hypergeometric*, where the m_{ij} behave like a set of binomial random variables conditional to M_j:

$$Pr(m_{1j};\ldots m_{k_j j}|M_j;n_{1j};\ldots n_{k_j j})=\left(M_j!(N_j-M_j)!\prod_{i=1}^{k} n_{ij}!\right) \Big/ \left(N_j!\prod_{i=1}^{k} m_{ij}!(n_{ij}-m_{ij})!\right)$$

When computed on an observed set of m_{ij}, this is the likelihood L_{j0} of observing the actual migrants class allocation for school j if H_0 holds. To define the denominator L^* for the LR test with a composite H_1, we simply assume that the observed outcome for migrants allocation is the generating model, i.e. that the probability to allocate a migrant children in class i of school j is m_{ij}/n_{ij}. When $k_j = 2$, the exact distribution of m_{1j} when allocation is not random is the *non-central* (or *extended*) *hypergeometric* (Zelterman, 2006), identified through its *odds ratio* parameter (the hypergeometric is a special case for $OR = 1$). OR will be estimated using the observed numbers of allocated students, as $\phi_j = [m_{1j}(n_{2j} - m_{2j})]/[m_{2j}(n_{1j} - m_{1j})]$. When $k_j > 2$ this leads to the *multivariate non-central hypergeometric distribution*, identified by the set of $k_j - 1$ independent *ORs* in the k_j x 2 table that express the ratio between the odds of a migrant of being allocated in class i instead of any other class, and the corresponding odds for a native. These will again be estimated using the observed frequencies as in $\phi_{ij} = \dfrac{m_{ij}}{M_j - m_{ij}} \cdot \dfrac{(N_j - M_j) - (n_{ij} - m_{ij})}{(n_{ij} - m_{ij})}$.

Let $\boldsymbol{m}_j$ be the vector of migrants allocated in each class of school j, and $\boldsymbol{n}_j$ the associated vector of class sizes. Then the above distribution can be written as:

$$Pr(\boldsymbol{m}_j | M_j;\ \boldsymbol{n}_j) = \left[\binom{M_j}{\boldsymbol{m}_j} \binom{N_j - M_j}{\boldsymbol{n}_j - \boldsymbol{m}_j} \prod_{i=1}^{k_j} \phi_i^{m_{ij}} \right] \Bigg/ \left[\sum_{\boldsymbol{h}} \binom{M_j}{\boldsymbol{h}_j} \binom{N_j - M_j}{\boldsymbol{n}_j - \boldsymbol{h}_j} \prod_{i=1}^{k_j} \phi_i^{m_{ij}} \right]$$

where the terms $\dbinom{M_j}{\boldsymbol{m}_j} = \dfrac{M_j!}{m_{1j}! m_{2j}! \ldots . m_{ij}!}$ are multinomial coefficients, and the denominator sums the terms for all possible allocation vectors $\boldsymbol{h}$. Again, total likelihood is $L^* = \prod_{j=1}^{S} L_j^*$. Thus, a test for H_0 can be defined for -2log LR with $\sum_{j=1}^{S} k_j - S$ degrees of freedom; since this is very large, the distribution under H_0 can be considered normal. Notice that LR can be written equivalently as $LR = \prod_{j=1}^{S} (L_{j0}/L_j^*) = \prod_{j=1}^{S} LR_j$.

For both tests, if H_0 is refused when evaluated for *all* schools, the largest subset of schools for which it holds can be determined repeatedly testing on the decreasing subsets of S-1, S-2,.... schools, until the global p-value approaches the desired α. For any subset size $r < S$, the r included schools should be selected testing all possible combinations and choosing the one with the highest p-value. Since this is computationally unfeasible, a simplified selection process is applied, excluding the schools with the highest p-value for the *single* school test (X^2_j or LR_j) one at a time and checking the global tests on the remaining subset. In this way, if the largest acceptable size is r, only $S - r$ tests will be performed. Finally, left out schools can undergo specific, more detailed analysis.

3 Results and discussion

For the first time, the 2010 Invalsi SNV survey included grades 2 and 5 (the one examined here) for all primary schools in Italy. Students were classified as *natives*, *first* or *second* generation migrants (the latter distinction is dropped here). Schools with no migrants were excluded, bringing the total number of involved primary schools to 5315.

Both the LR and the X^2 tests lead to a strong rejection of H_0 for the whole set of schools (*p-values* are very close to 0), implying that at least some schools do have some class allocation policy that tends to concentrate migrants children in a few classes. Notice that a similar X^2 on equal classroom allocation of males and females accepts the null hypothesis quite strongly, with a global X^2 sum of 16941 on 19938 degrees of freedom.

The selection procedure to find the largest subset of schools that evenly allocate students gave again similar results with the two tests. Sets defined by the X^2 test are slightly larger, but the difference w.r.t. to the whole system is limited (little more than 2%). Besides, the sets defined by the two tests have very large intersections (Table 2), and cases where the two tests disagree do not have lower expected frequencies (1,84 on average) than the others (1,71).

Table 1: Size of largest school subsets with H_0 accepted (α = 0,05 and 0,1)

	α = 0,1	% of total	α = 0,05	% of total
LR test	4381	82,92	4396	83,21
X^2 test	4528	85,19	4543	85,47

Table 2. Schools included in both *LR* and X^2 derived largest subsets

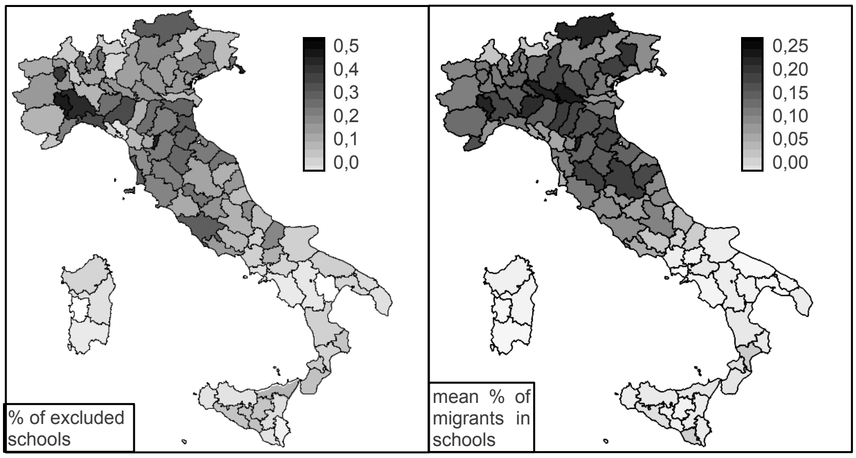
α = 0,1 intersection			α = 0,05 intersection		
size	% of *LR* subset	% of X^2 subset	size	% of *LR* subset	% of X^2 subset
4267	97,40	94,26	4287	97,52	94,36

Results for the two α are rather close, as α=0,1 adds to the set of random allocation schools less than 20 units more than α=0,05. This could be due to the large sample size, but could also suggest a rather neat partition of schools into the two sets, with allocating behavior that changes quite suddenly near the threshold between the two sets.

Note that the applied sequence of tests differs from the union of single school tests, defined by taking all schools where local *p*-values for each X^2_j or LR_j are smaller than α. The size r^* of the latter set would differ from the one determined by the global tests.

Fig. 1 shows on the left the spatial distribution of excluded schools to be quite uneven with lower levels in the South (where however migrants are fewer, as shown on the right side map) and values above 1/3 in a few North-West provinces. In general the mean proportion of migrants in excluded schools is higher (around 14%) than in schools with fair allocation policies (around 9%), but this holds for both North and South, though with different absolute levels (18% against 13% in the North and 7.5% against 4.5% in the South), so it does not seems to be a difference induced by different local approaches.

Figure 1: Schools excluded from the random allocation set and migrants per Province (*LR* test)



References

Contini D. (2011) Immigrant background peer effects in Italian schools, *Working paper*, Department of Economics, Università di Torino

Schnepf S. V. (2007) Immigrants' educational disadvantage, *J. of Population Economics* 20, 527-545

Zelterman D., 2006, *Models for discrete data*, Oxford University Press, Oxford.