University drop-out: The case of Italy

by Federico Cingano and Piero Cipollone
The purpose of the Temi di discussione series is to promote the circulation of working papers prepared within the Bank of Italy or presented in Bank seminars by outside economists with the aim of stimulating comments and suggestions.

The views expressed in the articles are those of the authors and do not involve the responsibility of the Bank.

Editorial Board: Domenico J. Marchetti, Marcello Bofondi, Michele Caivano, Stefano Iezzi, Andrea Lamorgese, Francesca Lotti, Marcello Pericoli, Massimo Sbracia, Alessandro Secchi, Pietro Tommasino.

Editorial Assistants: Roberto Marano, Alessandra Piccinini.
UNIVERSITY DROP-OUT: THE CASE OF ITALY

by Federico Cingano* and Piero Cipollone*

Abstract

We combine individual and aggregate-level data on educational attainment to study the determinants of university drop-out in Italy, one of the worst performers among developed countries. Based on detailed information on a representative sample of secondary school graduates and on local university supply we first show that family and educational background are relevant determinants of continuation probability. In particular, our results show that accounting for enrollment-induced sample selection significantly enhances the estimated coefficients with respect to standard probit analysis. We then combine our estimates with data on family and educational backgrounds of secondary school graduates in comparable European countries and find that differences in endowments only explain a minor fraction of the observed cross-country gap in students’ attainments.

JEL Classification: I21, I22, J62, C35.
Keywords: university drop-out, school transitions, social mobility, tobit estimation.

Contents

1. Introduction................................................................................................................ .......... 3
2. Determinants of University drop-out probability ................................................................ 5
   2.1 Statistical framework and empirical issues .................................................................. 5
   2.2 Data description........................................................................................................... .8
   2.3 Determinants of drop-out probability: results ............................................................ 11
3. Explaining cross-country attainment differences .............................................................. 13
4. Conclusions................................................................................................................. ....... 15
References .............................................................................................................................. 17
Tables ..................................................................................................................................... 19

* Bank of Italy, Economic Research Department.
1. Introduction

Both economists and sociologists have long been interested in the relationships between educational attainments and individual, familiar and environmental backgrounds (Mare, 1980; Willis and Rosen, 1979; Shavit and Blossfeld, 1993). Evidence suggesting strong dependence of educational outcomes from characteristics as gender, race or family conditions represents a relevant indicator of inequality in the opportunities of social mobility. In recent years policymakers and observers in many developed countries have focussed in particular on the low retention rate of tertiary education systems which might have increasingly negative distributional consequences given the widening college wage premium. Quite surprisingly, a recent strand of literature focussing on the determinants drop-out among university students in several countries found small or no role for individual variables as family or educational backgrounds (see Naylor and Smith, 2001; Johnes and McNabb, 2004 and Arulampalam et al 2001 for the UK; Montmarquette et al. 2001 for Canada; Jackobsen and Rosholm, 2003 for Denmark).

In this paper we study university withdrawal decisions in Italy, a country displaying one of the highest drop out rates among OECD members (58% against an average of 30%). Unlike the above mentioned papers, focusing on continuation decisions of enrolled university students, we base our analysis on a representative sample of Italian secondary school graduates. This allows controlling for selection biases arising when some determinants of the drop-out decision affects realization at previous transition (Cameron and Heckman, 1998; Keane and Wolpin, 1997). Our selection-corrected estimates exploit both functional forms and instrumental variables identification based on measures of anticipated costs of university attendance. Contrary to the existing evidence, our results point to a very relevant role of both family and educational background characteristics on continuation probabilities. For example,
we find that a ten years increase in father’s schooling (corresponding to moving from compulsory education to university degree) is associated to a reduction in dropout probability by 14 percentage points, against an average predicted probability of 22 percent. The estimated effects obtained not accounting selection are significantly smaller, nearly 5 percent, and in line with the above mentioned works.

We use these findings to check whether the disproportionately high university dropout rate in Italy can be explained in terms of the level and quality of schooling of the adult population. We combine individual data on parental backgrounds and educational curricula of secondary school graduates in four large European countries with our estimates and compute the fraction of the gap in observed drop-out rates attributable to differences in these variables. Despite the sometimes large cross-country gap in background variables, we find that assigning Italian parents the same average levels of education observed in comparable developed countries would in the best case scenario reduce withdrawal by just 6 percentage points. Changing the composition of secondary school graduates by type of school attended would not explain much of the gap either.3

While raising concerns on the effectiveness of the existing system of education in equalizing opportunities and promoting social mobility, our analysis thus suggests that changes in (observable) initial conditions are unlikely to yield significant reductions average withdrawal rates.

The rest of the paper is organized as follows. Section 2 illustrates the statistical framework we used in the analysis and discuss selection and endogeneity problems in the estimation; we subsequently describe the data set and present the results. Section 3 computes the cross-country comparative exercise. Section 4 briefly concludes.

---

3 Clearly, this exercise is “partial” since we can not take into account the potential effects of cross-country differences in endowments on continuation probabilities.
2. Determinants of University drop-out probability

2.1. Statistical framework and empirical issues

To illustrate the main selection issues involved in estimating the determinants of individual schooling attainment consider a simple statistical model assuming that the unobserved disutility associated to school attendance by individual $i$ ($y^*_i$), is determined according to:

$$ y^*_i = EB_i \beta_1 + FB_i \beta_2 + LC_i \beta_3 + X'_i \gamma + \varepsilon_i $$

In equation (1) $EB$ and $FB$ describe student $i$ educational and family background, respectively, $LC$ captures relevant local conditions, $X$ is a vector of individual characteristics and $\varepsilon_i$ is a disturbance term capturing residual unobserved heterogeneity. Students will drop-out if $y^*_i$ is higher than a given threshold, normalized to zero. Let $D$ be the drop-out indicator, then withdrawal ($D_i = 1$) is observed if $y^*_i > 0$. Dropout probability can therefore be written as:

$$ P(D_i = 1) = P(\varepsilon_i > -EB_i \beta_1 - FB_i \beta_2 - LC_i \beta_3 - X'_i \gamma). $$

When $\varepsilon_i$ distributes as a normal standard the above model can be estimated in a standard univariate Probit regression framework. This is the approach taken by recent studies on the determinants of university withdrawal (Naylor and Smith, 2001; Montmarquette et al. 2001). However, the simple occurrence that some variable affecting the choice to drop-out also determined outcomes at previous transitions implies sample selection bias would likely affect the estimated marginal probabilities.

To illustrate the nature of the distortion consider university enrollment decision and assume that familiar background (as parents’ education or income) is the only determinant of college enrollment ($E^*$) and drop out ($D^*$):

$$ D^*_i = FB_i \beta + \eta_i $$
$$ E^*_i = FB_i \alpha + \varepsilon_i $$
with $\alpha > 0$ and $\beta < 0$, respectively. Suppose that individual $i$ enrolls if $E_i^* > 0$ and drops out if $D_i^* > 0$ and assume for simplicity that $FB$ can be either 0 or 1. The average effect of $FB$ on withdrawal

$$ E(D_i^* \mid E_i^* > 0, FB_i = 1) - E(D_i^* \mid E_i^* > 0, FB_i = 0) = \beta + E(\eta_i \mid E_i^* > 0, FB_i = 1) - E(\eta_i \mid E_i^* > 0, FB_i = 0) $$

could be obtained estimating $\beta$ in the first equation only if:

$$ E(\eta_i \mid E_i^* > 0, FB_i = 1) - E(\eta_i \mid E_i^* > 0, FB_i = 0) = 0 $$

Since according to the selection equation the low-FB enrolled would necessarily have higher average unobservables than the high-FB enrolled, the condition above would not hold if $\text{corr}(\varepsilon, \eta) \neq 0$. For example, since higher draws from the distribution of $\varepsilon$ are required for students with bad as opposed to good family backgrounds to enroll, if $\text{corr}(\varepsilon, \eta) < 0$ they would also have lower chances to drop-out on average. If not accounted for, such selection mechanism would bias the estimated marginal effect of family background upwards in a single-probit regression of drop-out probability. Similar reasoning could apply to educational background and other relevant controls.

Enrollment-induced selection can be accounted for specifying a modified version of Tobit type 2 model (Amemiya 1985)

$$ D_i^* = X_{Di} \beta + \eta_i \tag{3} $$

$$ E_i^* = X_{Ei} \alpha + \varepsilon_i \tag{4} $$

where $D^*$ and $E^*$ are two latent variables representing, respectively, the propensity of each individual to enroll and subsequently withdraw and $X_{Di}$ and $X_{Ei}$ are different group of explanatory variables. In this framework, we only observe the sign of $E^*$ ($E^*>0$ indicating enrollment) and, when this is positive, the sign of $D^*$ ($D^*_i \leq 0$ indicating the student has not withdrawn). The following table summarizes the available information for this model:

<table>
<thead>
<tr>
<th>$E_i^* \leq 0$</th>
<th>$D_i^* \leq 0$</th>
<th>$D_i^* &gt; 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e_i=0$</td>
<td>$d_i=\text{unobserved}$</td>
<td>$e_i=0$, $d_i=\text{unobserved}$</td>
</tr>
<tr>
<td>$E_i^* &gt; 0$</td>
<td>$e_i=1$</td>
<td>$d_i=0$</td>
</tr>
<tr>
<td></td>
<td>$e_i=1$, $d_i=1$</td>
<td></td>
</tr>
</tbody>
</table>

where the couple $\{e_i, d_i\}$ represent the observed sample for individual $i$, $e_i$ is an indicator for college enrolment and $d_i$ is an indicator for drop out.
We also assume that \{\varepsilon_i, \eta_i\} are i.i.d. drawn from a bivariate distribution with zero mean, variances \(\sigma_1^2\) and \(\sigma_2^2\) and covariance \(\sigma_{12}\). The associate likelihood function for individual \(i\) would be:

\[
L_i = \left[ P(E_i^* \leq 0) \right]^{-e_i} \left[ P(D_i^* \leq 0 | e_i = 1) \right]^{-d_i} \left( P(D_i^* > 0 | e_i = 1) \right)^{d_i} \left( P(E_i^* > 0) \right)^{e_i}.
\]

The first part of the expression accounts for individuals who did not enroll, while the second takes care of university students that either dropped out (first term) or are still enrolled at the time of interview (second term). In order to estimate the above likelihood we assumed that the \{\varepsilon_i, \eta_i\} are jointly normal. This way our statistical framework represents a modified version of the Heckman selection model studied by Van de Ven and Van Praag (1981).

There are a variety of reasons why unobserved propensity to enroll and to continue tertiary studies might be positively correlated, implying that \(\text{corr}(\varepsilon_i, \eta_i) < 0\). One that has received considerable attention in the literature is “ability”. When not observed, differences in the cost each individual faces when acquiring education, either due to intellectual skills or motivation etc., are likely to induce severe biases. For example, our data show that nearly all (90%) children to academic father attend university, more than twice the share of students whose father only achieved compulsory education (see Table1). If ability is a relevant dimension for selection into University, enrolled students coming from more disadvantaged families would on average be more talented than their colleagues coming from richer families, attenuating the estimated impact of parents’ education (i.e. upward-bias).

When \(X_{Di} = X_{Ei}\) parameter identification in (5) simply rely on functional form assumptions. When the costs of attendance are an important component of enrollment decisions, however, one may exploit the identification power induced by individual-level variation in those costs. Indicators of the local supply of university courses, capturing the fact that students grown up in an area without college face higher costs of education, and/or the number of kids in the family, a proxy of the resources available per capita given household characteristics are two instruments used in the literature (Card, 1995; Cappellari, 2003). Relying on such sources of variation implies assuming that all effects of direct costs, affecting the expected private rate of returns to university education, are anticipated and included in the enrollment decision. Therefore implicit assumptions here would be that, conditional on socioeconomic characteristics (accounting for example for location choices) and early school
performances, drop-out decisions are determined by individual shocks (such as an update of their ability, motivation, tastes, etc.) that are unrelated to the local availability of University courses (and/or to family-size).

Given our sample of secondary school graduates is likely to be non-random, the model above will not allow us to recover population-parameters. Hence, our estimates would fail to predict the consequences of policies targeted at university drop-outs if such policies in turn affected the estimated coefficients through composition-effects on the sample of high-school graduates.

2.2. Data description

Our data originate from a survey realized in 2001 by the Italian National Statistical Institute (ISTAT) on nearly 23,000 individuals. The sample, consisting of approximately 5% of the population, is representative of students who got their secondary school degree in 1998, and contains very detailed information on their activity up to 2001, their educational background and both family and individual characteristics. The data allows in particular tracking the whole educational history of each individual, and provide a full description of academic or labor market performance during the three years after graduation at secondary schools. Furthermore it distinguishes between students currently enrolled at University, those who dropped out and those who entered the labor market.

More specifically, in our empirical analysis we will exploit the following information contained in the survey. Individual characteristics include sex, age, marriage, number of siblings and the place of residence –this data is available at a very detailed (i.e. municipality) geographical level. Indicators of past educational choices and performance are the degree obtained at the end of compulsory school (lower secondary school), the type of upper secondary school attended, the number of years taken to completion and the degree obtained. As to family background, while we do not have information on income, the data report both parents’ education (measured by years of formal education obtained when the student was 14), and parents’ profession (with a breakdown into entrepreneur, professional, high skilled

---

and low skilled white collar, blue collar, no qualification. We also know whether at least one of the grandparents had achieved higher education. We will combine such information with indicators of local conditions capturing spatial differences in the socio-economic environment that might be important determinants of educational outcomes. In particular we included the local unemployment rate in the place of residence and a measure of the degree of urbanization, captured by the population size of the municipality, both recovered from the National Population Census (ISTAT, 2001).

Table 3 presents students’ distributions according to secondary school attendance, and shows that the (weighted) sample provides a very good representation of the population along these two dimensions. According to our data more than 40% of students interviewed in 2001 had obtained a technical school degree in 1998, and almost a third of the sample attended General schools (“Licei”). These numbers are very closed to the population distribution. Figures in Table 3 indicate that the rate of response to the survey, conducted as a Computer Added Telephone Interview, has not significantly affected the sampling design as devised by the National Statistical Office. However graduates participating to the interviews might have tended to misreport their actual choices, in particular regarding University enrollment and dropout. As regards enrollment decisions, available administrative data allows comparisons with the population in terms of the ratio of students entering any tertiary education course in the same year they obtain the degree (see Tab.4). According to the Ministry of Education in 1998 this share (45.5%) was only slightly higher than the ratio provided by our sample (44.2%).

Discussing the sample representativeness in terms of dropout rates is slightly more complicated. Out of the 7483 students who entered tertiary education in 1998, 1048 declared to have given up studying within the three-year period covered by our survey. Unfortunately, there is no directly comparable administrative data reporting the dropout rate by cohort of graduates enrolled. One available proxy for the dropout rate is the share of students no longer enrolled in the same University course by year of enrollment. (Note that this measure, used as official dropout figure by the administration is likely to overestimate the abandonment rate since it includes students switching to a different course). In the academic year 2001/02 we find that, relative to students enrolled in 1998, such share amounted to 28% (Table 4). In our sample the share of 1998 graduates who left or changed university by summer of 2001
amounts to 23%, suggesting that our withdrawal rate is slightly underestimated. There are several reasons why the dropout rate turns out to be too low in the sample. First, dropouts could tend to misreport. Some of them could not even declare to have ever been enrolled, explaining the slightly lower share mentioned above. Others, though declaring to have enrolled, might not report the abandonment. We can attempt to control for such students by analyzing consistency of answers throughout the survey. For example one might think that students enrolled but having passed no exam within three years from enrollment are actual (or potential) dropouts. Including such students, the share of dropouts after three years from graduation in the sample rises to 25%. Also, graduate students who declared to have never been enrolled after graduation but reporting to have rejected some job offer or to have left a job to “better concentrate on their studies” could plausibly be imputed to the dropout population, as well those male who, after three years from graduation, have not yet joined the (compulsory) military service. In this case the share of dropouts rises to nearly 28%, in line with available administrative data.

As far as family background is considered, Table 5 show the sample distribution by degree completed by each parent, at the time students interviewed were 14 years old. In nearly 50% of cases both parents had at most completed compulsory education (8 years of formal schooling). Fathers are on average slightly more educated than mothers and assortative mating (i.e. families in which both parents tend to have the same amount of education) tend to prevail at low educational levels. Our data also indicate that the majority of secondary school graduates’ fathers were either blue collar or employed in the retail sector, while less than 20% were skilled or high skilled white collars (intellectual and scientific professions, qualified technicians, etc.).

---

5 Another common measure of the dropout rate is the complement to one of the “success rates”, obtained by comparing the number of University degrees obtained in a particular year with the number of students who enrolled some (in Italy, 7) years before. Estimated this way the drop out rate was 58.5% in 1997. A similar figure can be obtained in our sample grouping those who left any tertiary course with students still enrolled in 2001 who passed less than three exams per year.

6 Changing the outcome measure along these lines did not affect the results. Our preferred measure includes enrolled students who reported abandonment or declared to be employed full-time at the time of interview. On the other hand, including among drop-outs students who declared to be enrolled in a different course, as in the administrative definition, tends to attenuate the estimated effects.
2.3. Determinants of drop-out probability: results

In this section we discuss the main results from our University dropout probability model with selection, and compare them with univariate probit results. Following the literature on educational outcomes, we focus on a specification including indicators of family background, (both in terms of parents’ years of schooling, grandparents education and father’s profession), of past educational background and performance (the type of secondary school attended and the degrees obtained at the end of lower-primary, mandatory schools), and a set of individual variables (sex, age, marriage status). Controls for local conditions include local unemployment rate and degree of urbanization of the municipality the secondary school is located in. Summary statistics of the main variables used in the empirical part are presented in Table 6.

Column 1 in Table 7 reports the marginal effects on dropout probability of the main variables as estimated accounting for selection into university. To ease comparison with standard probit estimates (reported in column 3) we evaluate such effects setting all observable characteristics at the mean of the sub-sample of University enrolled. Results from our Tobit estimates indicate that both family background and educational background variables significantly affect withdrawal decisions. In particular, the drop-out probability is decreasing in father’s years of formal education: the estimated coefficient implies that a ten years increase in father’s schooling (corresponding to moving from compulsory education to University degree) reduces the dropout probability by 14 percentage points. Given the predicted probability at sample mean is 21%, the implied fall of withdrawal risk we estimate is considerable. As our discussion in section 2.1 suggested the effect obtained estimating a standard probit regression is substantially lower (5%). Similar conclusions can be drawn comparing the estimated coefficients on mother education.

How one should interpret the differences in educational responses by family background is a matter of debate in the recent literature on educational attainments (Card, 1999, 2001; Kane, 2001; Cameron and Heckman, 1998; Carneiro and Heckman, 2002). The main concurring explanations are short-term credit constraints and long-term factors fostering cognitive and non-cognitive abilities through a better learning environment or a higher quality of education. To discriminate between the two channels Cameron and Heckman (1998)
propose to estimate family-effects controlling for measures of early educational outcomes, which should absorb long-term factors. Our estimates of sizeable family-effects are obtained conditioning on the degree obtained at primary school: if interpreted in this framework, then, they would point to the existence of short-term credit constraints in education.

Our results point to a role for educational backgrounds, in that withdrawal probabilities decrease moving from Vocational to General schools. Again, accounting for selection magnifies the effect that would have been inferred without correction for the school-type effect on enrollment decisions, as a consequence of the fact that the same variables have exactly the opposite effect on enrollment than they have on withdrawal. Interpreting these coefficients is complicated by the fact that past educational choices might have induced sorting of students (for example, by learning abilities) into school types. To the extent that sorting based on learning abilities is accounted for by early educational outcomes, our results indicate that, for example, the predicted dropout probability for the average Vocational student would reduce by more than 50% if, other things equal, she had obtained a degree from a General school.

Finally, we find that female students have a lower dropout probability than their male colleagues. All other variables accounting for familiar background (grandparents’ education and father profession, not shown for brevity) and local conditions (as captured by the degrees of urbanization and rate of activity in the municipality) do not play any significant role. As far as University enrolment is concerned we find that, other things equal, enrollment probability increases substantially in the educational attainment of both parents, with almost identical coefficients. For example, the enrollment probability of children born to university graduates is 24% higher than it is for offspring of lower high school graduates. Conditional on parents’ education, enrollment is also strongly affected by the type of secondary school attended.

The availability of detailed individual information allowed us to test robustness of these findings to the use of instruments measuring tertiary education participation costs. In our exercise identification requires the anticipated costs of attendance determine the demand for education but do not directly affect outcomes once observable characteristics are taken

---

7 This would be the case if, for example, general schools attract all good students while all bad students choose “other” schools and yield upward bias estimates of school-type coefficients.
8 The reported marginal effects are evaluated at the secondary graduates (not just the enrolled) mean values of the observable characteristics.
into account. Using data from the Statistical Office of the Ministry of University and Research we constructed several measures of University courses availability at the local level. For every municipality in Italy we measured the distance from the nearest University and a distance-weighted index of university and degree subjects available in the entire territory. Although in Italy University tuition costs are generally low, large distances from Universities imply higher costs for households (in terms of transportation, rents etc.). As an additional measure of the actual availability of University courses we included the province-share of university enrolled in 1998 over the population aged 20-24. Second, we considered the number of siblings in the household. The larger the size of the family, the higher the probability that the observed student competes within the family (either due to scarce resources, or to the fact that each household attaches decreasing utility to one extra child enrolled, etc) and this lowers her enrollment probability, without affecting university outcomes.

The results obtained using the instrumental variables described above are reported in Table 8. In columns 1 and 2 we considered changes in the cost of enrollment induced by geographic variation in the availability of tertiary education courses, while in columns 3 and 4 we also accounted for the effect of different family sizes. The estimated coefficients confirm the relevance of accounting for enrollment decisions in studies on the effects of socio-economics status and educational background.9

3. Explaining cross-country attainment differences

All commonly used indicators of educational attainment point to the existence of large differences in completion rates between Italy and comparable European countries. According to the OECD, for example, non-completion rates in Europe ranged from about 20 per cent in the United Kingdom and Ireland to 40 per cent in Austria and France, and reached nearly 60 per cent in Italy (OECD, 2003). While these figures might represent country-specific

---

9 Appropriate measurement of the drop-out indicator seems to be also important for the results. When including among drop-outs students who changed university course within the period considered, a definition that is closer to the administrative data on withdrawal, both estimates of family and educational background coefficients became considerably weaker. This suggests that the use of data assembled by single universities with no possibility to control for spurious withdrawals (as in Montmarquette et al. 2000) could further bias the inference against the existence of family and educational background effects.
equilibrium outcomes if the population of university enrolled differed in unobservable individual characteristics as motivation, discount rates etc. (see Eckstein and Wolpin, 1999), our results allow us to evaluate the relevance of an alternative explanation based on differences in observables.

The first column in Table 9 reports the average years of formal education accumulated by parents of secondary school graduates in several European countries, computed from the 1998 issue of the European Community Households Panel (ECHP). Figures for Italy are very close to those we obtained in our sample and are lower than those of other countries. The second column reports the existing differences in the secondary school graduates as to the type of school (“program orientation”) attended according to OECD statistics: Italian graduates from general schools, associated to higher survivor probability than vocational schools, are fewer than in comparable European countries. In Table 10 we report the changes in the OECD figures for drop-out rate (defined as the share of university enrolled having abandoned before the fifth year) and survivor rates (the ratio of survivors at the fifth year relative to the population in the relevant age cohort), computed combining these data with our estimates. Specifically, column 1 shows the drop-out rate obtained attributing Italian secondary school graduates foreign family backgrounds as reported in table 10. The estimated reduction in withdrawal ranges from 2 to 9.3 percent, leaving the average observed gap above 20 percentage points. In column 2 we apply a similar procedure to calculate the effect of a change in the ratio of secondary school graduates from general schools to population (correspondingly lowering the shares in other schools) to other countries’ level. The implied reduction in withdrawal rate ranges from 0.5 to less than 5 percentage points. Combining both exercises would induce reductions of the drop-out rates ranging from about 3 to 10.3%. The difference in withdrawal rate would in the best case scenario (France) still be as large as 19%. The impact on the survivor rate is computed similarly but account for the effects of changes in the observable variables on enrollment rates. Results showed in columns 5 to 7 indicate the ratio of fifth-year enrolled students to population could in the best case

---

10 Given our estimates refer to the continuation probability at the third year, while OECD drop out and survivor rates are computed at the fifth year the estimated effect had to be extrapolated. Details are reported in the note to the table.
(UK, both exercises combined) increase by 9 percentage points. It would be still more than 30 points below the actual survivor rate in the United Kingdom, however.\textsuperscript{11}

We are unfortunately unable to assess to what extent changes in the sample composition as to family and educational backgrounds could affect the simulated attainments rates through changes in the estimated coefficients. However, the magnitude of the unexplained differential in attainments suggests the role played by country-specific characteristics in explaining “productivity” differences should be extremely relevant.\textsuperscript{12} One such characteristic, one that motivated recent reforms of the university system, is the limited supply of tertiary Type-B (i.e. three year) courses with respect to other countries.\textsuperscript{13} Alternatively, the population of Italian students might differ as to unobservable characteristics affecting the opportunity cost of attendance, as motivation, discount rates or outside opportunities. In this case a higher fraction of the enrolled would be willing to withdraw as they receive offers from the labor market. Finally, Italian students might be relatively more.

4. Conclusions

We exploit a survey conducted on a representative sample of Italian high school graduates to study the determinants of university drop-out accounting for enrollment-induced selection. Contrary to recent empirical work focusing on samples of university enrolled, our results indicate differences in background individual characteristics, and in particular in family characteristics, play a determinant role in explaining withdrawal. Comparing our results with those obtained with standard univariate analysis allowed us to determine the bias

\textsuperscript{11} Since in our estimations survivor rates are obtained relative to the population of secondary school graduates, as opposed to the entire population cohort used by OECD, here we need to assume that the ratio of Secondary School graduates to the relevant population cohort is constant, that is it does not change with family background. An alternative would be to report the simulated ratios of Survivors to Secondary School graduates, but this is not a commonly used OECD figure.

\textsuperscript{12} Interestingly, recent estimates of the returns to investing in tertiary education by Ciccone et al. (2004) suggest that differences in the expected economic gains from university attendance should not play a major role either. In year 2000 the private returns to university education, calculated as the discount rate that equates the present value of the additional costs of attendance to the present value of the stream of net-of-tax earnings generated by an increase in education, in Italy was above 10%, broadly in line with returns in Germany, France and Spain (and much larger than the returns to alternative investment).

\textsuperscript{13} Note, however, that the drop-out rate in Italy is not much lower in short- than long-term courses (49 as opposed to 58 per cent according to OECD data). Simple calculations obtained redistributing students across ISCED type A and B courses as in reference countries show the drop out rate would reduce by at most 3.4%.
induced by sample selection is substantial. Despite being large, however, background conditions do not seem to be able to explain why the drop-out rate in Italy is so higher than in comparable countries.

In terms of policy, our analysis confirms the strong concerns regarding the ability of the Italian educational system to promote social cohesion via equal educational opportunities. It suggests the role of familiar background mainly reflects short-term financial constraints rather than long term effect shaping offspring ability at early ages. Finally, it indicates that large differences in university completion rates might persist with respect to other countries even if educational attainments in the population converged. We are unfortunately unable to assess the relevance of alternative explanations including differences in individual unobservables determining students attachment, which would require increasing selectivity (raising tuitions, selection at entry, etc.), or higher exposure to adverse unanticipated shocks due to lower access to (either public or private) credit.
References


OECD (2003), Education at glance. OECD Indicators 2002, Paris, France


Table 1

**Enrollment rates by father schooling**
(percentage points)

<table>
<thead>
<tr>
<th>Father schooling</th>
<th>No degree (5 years)</th>
<th>Primary school (8 years)</th>
<th>Junior high school (10 years)</th>
<th>Professional diploma (10 years)</th>
<th>High school (13 years)</th>
<th>College (18 years)</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not enrolled</td>
<td>83</td>
<td>67</td>
<td>57</td>
<td>54</td>
<td>33</td>
<td>11</td>
<td>48</td>
</tr>
<tr>
<td>Enrolled</td>
<td>17</td>
<td>33</td>
<td>43</td>
<td>46</td>
<td>67</td>
<td>89</td>
<td>52</td>
</tr>
<tr>
<td>Total</td>
<td>1</td>
<td>19</td>
<td>37</td>
<td>5</td>
<td>29</td>
<td>9</td>
<td>100</td>
</tr>
</tbody>
</table>


Table 2

**Enrollment rates by type of secondary school**
(percentage points)

<table>
<thead>
<tr>
<th>Type of school</th>
<th>Enrolled</th>
<th>Not Enrolled</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vocational schools</td>
<td>20</td>
<td>80</td>
<td>15</td>
</tr>
<tr>
<td>Technical schools</td>
<td>36</td>
<td>64</td>
<td>42</td>
</tr>
<tr>
<td>Other schools</td>
<td>49</td>
<td>51</td>
<td>13</td>
</tr>
<tr>
<td>Licei</td>
<td>93</td>
<td>7</td>
<td>30</td>
</tr>
<tr>
<td>TOTAL</td>
<td>52</td>
<td>48</td>
<td>100</td>
</tr>
</tbody>
</table>

### Table 3

**Distribution of population and (weighted) sample by type of secondary school**

<table>
<thead>
<tr>
<th></th>
<th>POPULATION</th>
<th>SAMPLE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>Frequency</td>
</tr>
<tr>
<td>TOTAL</td>
<td>485,150</td>
<td></td>
</tr>
<tr>
<td>Vocational schools</td>
<td>74,016</td>
<td>15.3</td>
</tr>
<tr>
<td>Technical schools</td>
<td>207,398</td>
<td>42.7</td>
</tr>
<tr>
<td>Licei</td>
<td>141,759</td>
<td>29.2</td>
</tr>
<tr>
<td>Other schools</td>
<td>61,977</td>
<td>12.8</td>
</tr>
</tbody>
</table>


### Table 4

**Comparison between population and (weighted) sample**  
(percentage points)

<table>
<thead>
<tr>
<th></th>
<th>POPULATION</th>
<th>SAMPLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enrolled/graduates in 1998</td>
<td>45.5</td>
<td>44.2</td>
</tr>
<tr>
<td>Dropouts/enrolled in 2001</td>
<td>28</td>
<td>23</td>
</tr>
</tbody>
</table>

Table 5

Formal education of parents
(percentage points)

<table>
<thead>
<tr>
<th>Years of formal education:</th>
<th>Mother</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No degree</td>
</tr>
<tr>
<td>Father</td>
<td></td>
</tr>
<tr>
<td>No degree</td>
<td>0.3</td>
</tr>
<tr>
<td>Primary school</td>
<td>0.3</td>
</tr>
<tr>
<td>Lower high school</td>
<td>0.1</td>
</tr>
<tr>
<td>Professional Diploma</td>
<td>0.0</td>
</tr>
<tr>
<td>High school</td>
<td>0.0</td>
</tr>
<tr>
<td>University</td>
<td>0.0</td>
</tr>
<tr>
<td>All</td>
<td>0.7</td>
</tr>
</tbody>
</table>


Note: Primary school implies 5 years of schooling, lower high school 3, Professional Diploma 2 or 3 and it is an alternative to High school (5 years). Finally the university degree requires 5 or 6 years. Compulsory schooling amounts to 8 years.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>22.9</td>
<td>3.48</td>
</tr>
<tr>
<td>Female</td>
<td>.53</td>
<td>.49</td>
</tr>
<tr>
<td>Father’s years of schooling</td>
<td>9.86</td>
<td>3.92</td>
</tr>
<tr>
<td>Mother’s years of schooling</td>
<td>9.45</td>
<td>3.82</td>
</tr>
<tr>
<td>Grandparents holding a college degree?</td>
<td>.11</td>
<td>.313</td>
</tr>
<tr>
<td>Father occupation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entrepreneur</td>
<td>.082</td>
<td>.275</td>
</tr>
<tr>
<td>High skilled</td>
<td>.086</td>
<td>.281</td>
</tr>
<tr>
<td>White Collar</td>
<td>.390</td>
<td>.487</td>
</tr>
<tr>
<td>Blue Collar</td>
<td>.379</td>
<td>.485</td>
</tr>
<tr>
<td>No qualification</td>
<td>.068</td>
<td>.252</td>
</tr>
<tr>
<td>Number of siblings</td>
<td>1.32</td>
<td>.902</td>
</tr>
<tr>
<td>Mark in Junior high school</td>
<td>2.33</td>
<td>1.18</td>
</tr>
<tr>
<td>Mark in high school</td>
<td>45.24</td>
<td>7.15</td>
</tr>
<tr>
<td>Repeated some classes</td>
<td>.215</td>
<td>.410</td>
</tr>
<tr>
<td>Unemployment rate</td>
<td>.203</td>
<td>.120</td>
</tr>
<tr>
<td>North-west</td>
<td>.208</td>
<td>.406</td>
</tr>
<tr>
<td>North-east</td>
<td>.153</td>
<td>.360</td>
</tr>
<tr>
<td>Center</td>
<td>.197</td>
<td>.398</td>
</tr>
<tr>
<td>South</td>
<td>.440</td>
<td>.496</td>
</tr>
</tbody>
</table>

The effect of family background, school choice and ability on University drop-out probability

<table>
<thead>
<tr>
<th></th>
<th>Probit with selection</th>
<th>Probit 2</th>
<th>Probit 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Marginal effects on</td>
<td>Marginal effects on</td>
<td>Marginal effects on</td>
</tr>
<tr>
<td></td>
<td>Drop Out</td>
<td>Enrolment</td>
<td>Drop Out</td>
</tr>
<tr>
<td></td>
<td>coeff.</td>
<td>s.e.</td>
<td>coeff.</td>
</tr>
<tr>
<td>Family background</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Father schooling</td>
<td>-0.0145</td>
<td>0.0022</td>
<td>0.0143</td>
</tr>
<tr>
<td>Mother schooling</td>
<td>-0.0070</td>
<td>0.0021</td>
<td>0.0106</td>
</tr>
<tr>
<td>High school variables</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General schools</td>
<td>-0.5630</td>
<td>0.0196</td>
<td>0.6738</td>
</tr>
<tr>
<td>Technical schools</td>
<td>-0.2348</td>
<td>0.0246</td>
<td>0.2531</td>
</tr>
<tr>
<td>Arts and teaching</td>
<td>-0.1101</td>
<td>0.0137</td>
<td>0.1212</td>
</tr>
<tr>
<td>Vocational schools</td>
<td>Reference</td>
<td></td>
<td>Reference</td>
</tr>
<tr>
<td>Past educational outcomes</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>D grade at junior H.S.</td>
<td>-0.0729</td>
<td>0.0157</td>
<td>0.0604</td>
</tr>
<tr>
<td>C grade at junior H.S.</td>
<td>-0.1522</td>
<td>0.0184</td>
<td>0.1738</td>
</tr>
<tr>
<td>B grade at junior H.S.</td>
<td>-0.2250</td>
<td>0.0204</td>
<td>0.2480</td>
</tr>
<tr>
<td>A grade at junior H.S.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>20476</td>
<td></td>
<td>8280</td>
</tr>
</tbody>
</table>

NOTE: Maximum likelihood estimation. Marginal Effects are computed evaluating the density function at the corresponding sample mean. The estimated specification includes controls for individual gender and age, for father profession (entrepreneur, professional, high skilled and low skilled white collar, blue collar, no qualification), mother employment status and qualification, grandparents education, and for local unemployment rate.
### Table 8

The effect of family background and school choice on University drop-out probability. IV estimates

<table>
<thead>
<tr>
<th></th>
<th>IV Supply</th>
<th>IV Supply + Family size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Marginal effects on Drop Out</td>
<td>Marginal effects on Enrolment</td>
</tr>
<tr>
<td></td>
<td>coeff.</td>
<td>s.e.</td>
</tr>
<tr>
<td><strong>Family background</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Father schooling</td>
<td>-0.0130</td>
<td>0.0036</td>
</tr>
<tr>
<td>Mother schooling</td>
<td>-0.0028</td>
<td>0.0028</td>
</tr>
<tr>
<td><strong>High school variables</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General schools</td>
<td>-0.4108</td>
<td>0.1039</td>
</tr>
<tr>
<td>Technical schools</td>
<td>-0.1923</td>
<td>0.0488</td>
</tr>
<tr>
<td>Arts and teaching</td>
<td>-0.0857</td>
<td>0.0251</td>
</tr>
<tr>
<td>Vocational schools</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td><strong>Past educational outcomes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D grade at junior H.S</td>
<td>-0.0828</td>
<td>0.0194</td>
</tr>
<tr>
<td>C grade at junior H.S</td>
<td>-0.1192</td>
<td>0.0325</td>
</tr>
<tr>
<td>B grade at junior H.S</td>
<td>-0.1907</td>
<td>0.0407</td>
</tr>
<tr>
<td>A grade at junior H.S</td>
<td>-0.0006</td>
<td>0.0003</td>
</tr>
<tr>
<td><strong>Instruments</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance</td>
<td>-0.0006</td>
<td>0.0003</td>
</tr>
<tr>
<td>0 siblings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 siblings</td>
<td>-0.0480</td>
<td>0.0193</td>
</tr>
<tr>
<td>2 siblings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 siblings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;4 siblings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ratio enrolled</td>
<td>0.1989</td>
<td>0.0943</td>
</tr>
<tr>
<td><strong>Observations</strong></td>
<td>20476</td>
<td></td>
</tr>
</tbody>
</table>

**NOTE:** Maximum likelihood estimation. Marginal Effects are computed evaluating the density function at the corresponding sample mean. The estimated specification includes controls for individual gender and age, for father profession (entrepreneur, professional, high skilled and low skilled white collar, blue collar, no qualification), mother employment status and qualification, grandparents education, and for local unemployment rate. Distances measured in km.
Table 9

Familiar and educational backgrounds of secondary school graduates in some European countries

<table>
<thead>
<tr>
<th>Father</th>
<th>Mother</th>
<th>General school</th>
</tr>
</thead>
<tbody>
<tr>
<td>Germany</td>
<td>13.2</td>
<td>5.7</td>
</tr>
<tr>
<td>UK</td>
<td>12.7</td>
<td>10.5</td>
</tr>
<tr>
<td>Spain</td>
<td>10.0</td>
<td>7.6</td>
</tr>
<tr>
<td>France</td>
<td>9.9</td>
<td>8.8</td>
</tr>
<tr>
<td>Italy</td>
<td><strong>9.8</strong></td>
<td><strong>5.8</strong></td>
</tr>
</tbody>
</table>


Table 10

Simulated Drop-out and survivor rates in Italy

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drop out rate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FB</td>
<td>51.8</td>
<td>56.8</td>
<td>50.8</td>
<td>28.5</td>
<td>22.9</td>
<td>19.2</td>
<td>24.3</td>
<td>30.8</td>
</tr>
<tr>
<td>EB</td>
<td>48.5</td>
<td>56.9</td>
<td>47.5</td>
<td>17.0</td>
<td>26.2</td>
<td>19.7</td>
<td>27.6</td>
<td>61.4</td>
</tr>
<tr>
<td>Both</td>
<td>55.8</td>
<td>53.5</td>
<td>51.5</td>
<td>23.7</td>
<td>20.2</td>
<td>24.1</td>
<td>26.0</td>
<td>48.1</td>
</tr>
<tr>
<td></td>
<td>55.0</td>
<td>57.3</td>
<td>54.5</td>
<td>36.3</td>
<td>21.0</td>
<td>19.2</td>
<td>21.6</td>
<td>37.0</td>
</tr>
<tr>
<td>Italy</td>
<td>57.8</td>
<td>57.8</td>
<td>57.8</td>
<td>57.8</td>
<td>18.6</td>
<td>18.6</td>
<td>18.6</td>
<td>18.6</td>
</tr>
</tbody>
</table>

Note: Following OECD definitions, the drop out rate is the ratio of university leavers to the relevant cohort of enrolled five years before; the survival rate is the ratio of students enrolled in the last (fifth) year to the relevant population cohort. The reported coefficients are obtained assuming the yearly drop-out rate (σ) is constant and writing the survivor probability of enrolled students at year \( t \) as \( S(t) = 1 - D(t) = \sigma t \), where \( D(t) \) is the corresponding drop-out rate. OECD figures refer to probabilities computed at the fifth year of enrollment \( (t=5) \) while our estimates refer to the third year. To simulate the drop-out rates reported in cols. 1 to 4 we proceeded as follows. First we computed \( \sigma_{\text{hat}} = (1-0.578)^{1/5} \), the yearly survivor rate implied by the OECD drop-out figure for Italy \( (D(t) = 0.578) \). We then simulated the change in survivor probability at third year \( (\sigma') \) multiplying the change in each of the explanatory variables indicated (family and educational background) with respect to country A \((A'X)\) by the corresponding coefficients estimated in table 7 \((A'\sigma_{\text{hat}}') = \beta A'X\). The simulated third-year survivor rate is therefore \( \sigma_{\text{hat},A} = \sigma_{\text{hat}} + A'\sigma_{\text{hat}}' \) which we project at the fifth year to obtain \( D_A(5) = 1 - \sigma_{\text{hat},A} \), reported in the table. The results for survival rates (cols. 5 to 7), defined by OECD with respect to the relevant population cohort, are computed similarly, and account for the effects of changes in \( X \) on enrollment rates. Observed ratios in the last row and in col. 4 and 8 are obtained from OECD “Education at glance. OECD Indicators 2002”.

25
RECENTLY PUBLISHED “TEMI” (*)

N. 602 – Scenario based principal component value-at-risk: An application to Italian banks’ interest rate risk exposure, by Roberta Fiori and Simonetta Iannotti (September 2006).
N. 603 – A dual-regime utility model for poverty analysis, by Claudia Biancotti (September 2006).
N. 604 – The political economy of investor protection, by Pietro Tommasino (December 2006).
N. 605 – Search in thick markets: Evidence from Italy, by Sabrina Di Addario (December 2006).
N. 606 – The transmission of monetary policy shocks from the US to the euro area, by S. Neri and A. Nobili (December 2006).
N. 608 – Merge and compete: Strategic incentives for vertical integration, by Filippo Vergara Caffarelli (December 2006).
N. 609 – Real-time determinants of fiscal policies in the euro area: Fiscal rules, cyclical conditions and elections, by Roberto Golinelli and Sandro Momigliano (December 2006).
N. 610 – L’under-reporting della ricchezza finanziaria nell’indagine sui bilanci delle famiglie, by Leandro D’Aurizio, Ivan Faiella, Stefano Iezzi, Andrea Neri (December 2006).
N. 611 – La polarizzazione territoriale del prodotto pro capite: un’analisi del caso italiano sulla base di dati provinciali by Stefano Iezzi (December 2006).
N. 612 – A neural network architecture for data editing in the Bank of Italy’s business surveys by Claudia Biancotti, Leandro D’Aurizio and Raffaele Tartaglia-Polcini (February 2007).
N. 613 – Outward FDI and Local Employment Growth in Italy, by Stefano Federico and Gaetano Alfredo Minerva (February 2007).
N. 614 – Testing for trend, by Fabio Busetti and Andrew Harvey (February 2007).
N. 615 – Macroeconomic uncertainty and banks’ lending decisions: The case of Italy, by Mario Quagliariello (February 2007).
N. 616 – Entry barriers in italian retail trade, by Fabiano Schivardi and Eliana Viviano (February 2007).
N. 617 – A policy-sensible core-inflation measure for the euro area, by Stefano Siviero and Giovanni Veronese (February 2007).
N. 618 – Le opinioni degli italiani sull’evasione fiscale, by Luigi Cannari and Giovanni D’Alessio (February 2007)
N. 619 – Memory for prices and the euro cash changeover: An analysis for cinema prices in Italy, by Vincenzo Cestari, Paolo Del Giovane and Clelia Rossi-Arnaud (February 2007).
N. 620 – Intertemporal consumption choices, transaction costs and limited participation in financial markets: Reconciling data and theory, by Orazio P. Attanasio and Monica Paiella (April 2007).
N. 624 – Un nuovo metodo per misurare la dotazione territoriale di infrastrutture di trasporto, by Giovanna Messina (April 2007).
N. 625 – The forgone gains of incomplete portfolios, by Monica Paiella (April 2007).

(*) Requests for copies should be sent to:


2002


E. GAOITTI and A. GENERALE, *Does monetary policy have asymmetric effects? A look at the investment decisions of Italian firms*, Giornale degli economisti e annali di economia, v. 61, 1, pp. 29-60, TD No. 429 (December 2001).


2003


2004


F. SPADAFORA, Il pilastro privato del sistema previdenziale: il caso del Regno Unito, Economia Pubblica, 34, (5), pp. 75-114, TD No. 503 (June 2004).


P. Del Giovane and R. Sabbatini, La divergenza tra inflazione rilevata e percepita in Italia, Bologna, Il Mulino, TD No. 532 (December 2004).

R. Torrini, Quota dei profitti e redditività del capitale in Italia: un tentativo di interpretazione, Politica economica, v. 21, pp. 7-42, TD No. 551 (June 2005).


L. Casolaro, L. Gambacorta and L. Guiso, Regulation, formal and informal enforcement and the development of the household loan market. Lessons from Italy, in Bertola G., Grant C. and Disney R. (eds.) The Economics of Consumer Credit: European Experience and Lessons from the US, Boston, MIT Press, TD No. 560 (September 2005).


2006


G. M. TOMAT, Prices product differentiation and quality measurement: A comparison between hedonic and matched model methods, Research in Economics, No. 60, pp. 54-68, TD No. 547 (February 2005).


2007


FORTHCOMING


F. LIPPI and S. NERI, Information variables for monetary policy in a small structural model of the euro area, Journal of Monetary Economics TD No. 511 (July 2004).


R. BRONZINI and G. DE BLASIO, Una valutazione degli incentivi pubblici agli investimenti, Rivista Italiana degli Economisti, TD No. 582 (March 2006).