
Working Paper Series

09/16

SOCIAL SELECTION IN HIGHER EDUCATION. ENROLMENT, DROPOUT AND TIMELY DEGREE ATTAINMENT IN ITALY.

**DALIT CONTINI, FEDERICA CUGNATA and
ANDREA SCAGNI**

Social selection in higher education. Enrolment, dropout and timely degree attainment in Italy.

Dalit Contini*, Federica Cugnata**, Andrea Scagni*

* University of Torino

** University Vita-Salute San Raffaele

Abstract

In this paper, we provide a picture of social selection throughout higher education in Italy, analysing a retrospective survey held in 2011 on the cohort of high school graduates 2007. We study enrolment, university system dropout and timely completion. Firstly, we model each outcome with separate logistic regressions, to examine the direct and indirect role of socioeconomic background via prior schooling. Secondly, we jointly analyse these results: by plotting the estimates of the retention probability (given enrolment) against the enrolment probability for subgroups of children by socio-demographic characteristics and prior schooling, we visualize the degree to which the disadvantage related to university enrolment also relates to retention, and acknowledge the existence of impressive inequalities. Thirdly, we jointly analyse retention and timely completion, and find that these two outcomes are affected differently by individual factors. Lastly, we examine the role of labour market conditions on higher education outcomes at the onset of the recent economic crisis: youth unemployment rates were negatively related to enrolment, timely completion and retention. The negative relation with retention suggests that there is little evidence in favour of explanations of dropout referring to labour market “pulling out” students from the university system.

1. Introduction

The reduction of inequalities in access to the highest levels of education and the increase in the number of young people with higher education degrees are strategic targets of the EU. In this perspective, there is a need to foster both enrolment decisions and retention in higher education, in particular for young people of low socioeconomic origin. Despite the rising participation rates occurred in the last decades, the share of individuals with a higher education qualification is still low in Italy as compared to the majority of EU countries (OECD, 2013). Moreover, socioeconomic inequalities in educational choices are very large. The relevant share of children not attaining the upper secondary school diploma, low university entry rates and high university dropout rates can be largely ascribed to the low educational attainment of children from

disadvantaged backgrounds. Given the highly stratified character of the Italian secondary school system, a crucial stage is the transition to upper secondary education. As shown in Jackson, Jonsson (2013), the role of prior school performance in shaping upper secondary school choices is similar to other countries, but socioeconomic differentials given performance are especially large in Italy. Hence, research on socioeconomic differentials in higher education should acknowledge the different selection processes to which students of different backgrounds are exposed. Regrettably, the study of educational inequalities in Italy is undermined by the lack of longitudinal data on children's schooling progress, so it is difficult to provide a comprehensive account of social selectivity from childhood to higher education.

Previous literature on higher education inequalities focused on enrolment and early dropout; instead, due to lack of data, research on degree completion is scant. By exploiting the most recent available wave of the Survey on High-school Graduates (held in 2011 on students completing upper secondary school in 2007), in this paper we analyse higher education *enrolment*, retention vs. *dropout* from the university system within 4 years from enrolment, updating the existing evidence based on earlier editions of the survey. In addition, for those attending a 3-year first-level degree, we also analyse *timely completion* (defined as degree attainment within 4 years from enrolment). To our knowledge, this is the first attempt to study timely completion in Italy.¹

In addition, we contribute to the existing literature in two directions. Firstly, after analysing social background differentials in each outcome, we relate these findings to provide new descriptive evidence of the socially selective process through higher education. We find that disadvantage cumulates over time: social groups who are disadvantaged in terms of enrolment are also disadvantaged in terms of retention and timely completion. When we look at the entire picture, we acknowledge the existence of dramatic inequalities, even between individuals with very similar prior schooling.

Secondly, we analyse the relation between labour market conditions and higher education outcomes. Merging the survey data with aggregate data at the provincial level, we exploit the territorial variation of youth unemployment rates at the time of graduation and analyse the relation between labour market conditions and higher education outcomes at the onset of the ongoing economic crisis. This exercise allows shedding some light on geographical inequalities: our results indicate that in high unemployment areas

¹ In previous waves of the survey, respondents were interviewed 3 years after the end of high-school, but only few students graduate perfectly in time.

enrolment, retention and timely completion probabilities are lower, suggesting that when labour market prospects are very poor, discouragement attitudes prevail.

2. Background

A common explanation of socioeconomic differences in educational choices in the sociological literature refers to the theory of cultural reproduction (Bourdieu, Passeron, 1990), according to which children in the highest classes are advantaged in gaining educational credentials due to the possession of cultural capital. Rational action (Breen, Golthorpe, 1997) is a competing explanation. In this view, individuals choose among educational options by evaluating costs and benefits and the perceived probabilities of successful outcomes. Individuals aim at reducing the risk of downward intergenerational mobility, so given performance, higher-class children make more ambitious educational investments. These mechanisms provide a rational choice account of the evidence that social class differentials operate *via* performance and *net* of performance (Boudon 1974). Clearly, performance differentials at the transition to higher education are originated also by social selection occurred at earlier stages of schooling. At the international level, there is evidence of substantial inequalities in upper secondary and tertiary education enrolment choices. However, cross-country differences are wide, and Italy stands out as a country with relatively large socioeconomic inequalities, overall and net of prior performance (Jackson, 2013).

Economists' explanations of educational inequalities also refer to rational choices: individuals take decisions by comparing direct and indirect costs of education, in particular, tuition fees and foregone earnings, with benefits in terms of future wages (e.g. Blundell *et al.*, 2001). Lower income individuals make less prestigious choices because they are more risk averse (for evidence on Italy, see Checchi *et al.*, 2014) or because of credit constraints. The role of credit constraints and financial hardship on enrolment decisions was addressed in particular in the USA, with mixed results. Carneiro, Heckman, (2002) find that what matters is not current income, but the long-run factors associated with higher income families, providing better quality education and better environments that foster cognitive and non-cognitive skills.

There is widespread evidence of socioeconomic differentials also in retention and completion probabilities. In first place, students from high social background are advantaged because they have better prior academic preparation. Tinto (1975) identifies students' academic and social integration as major determinants of completion in higher education. Robin and Naylor (2001) focus on academic preparedness, and

Arulampalam *et al.* (2004) highlight the role of individual's prior performance relative to that of the other students. Stinebrickner, Stinebrickner (2009, 2013), show that students update their beliefs on their own ability over time and this process plays a role in dropout decisions. However, inequalities may exist also when comparing individuals with similar schooling history and prior performance. Vignoles, Powdthavee (2009) find that, even after controlling for personal characteristics, prior achievement and university features, dropout probabilities in UK are larger for low social strata. Similarly, Ishitani (2006) finds that first-generation college students in the USA face higher dropout and lower timely completion probabilities than students with at least one parent with tertiary education. Other potential explanations are related to information asymmetries on the higher education system (if higher background children make better-informed choices, they will face lower risk of withdrawal), or to credit constraints (Stinebrickner, Stinebrickner 2008), tuition fees and financial aid (Dynarski 2003). Supply side factors may also matter, given that the choice of the institution depends on family background, since dropout and timely completion probabilities vary with institutions' characteristics (Bound *et al.*, 2009; Kurlaender *et al.* 2014).

Evidence on Italy

Research on enrolment in Italy on the demand side, largely focuses on intergenerational transmission of educational attainment over time (e.g. Checchi *et al.* 2013), and on the effect of the reform of the higher education system implemented in 2001, (Cappellari, Lucifora, 2009; Bratti *et al.*, 2010). On the supply side, the institutions' quality explains to some extent the geographical mobility occurring from South to North (Ciriaci, 2013; Pigni, Staffolani, 2015). There is also some evidence that the distance from the nearest institution and the variety of degree programs in near institutions affect enrolment decisions. Yet, enrolment rates do not vary much across macro-areas, despite the different density of institutions over the territory.

Research on university dropout suffers from the unavailability of longitudinal data at the national level.² Some studies rely on administrative data from specific departments or degree-programs in given institutions. However these data do not allow distinguishing between change of degree-program and withdrawal from higher education altogether. The data mostly used to study system-level dropout is the Survey of High-School Graduates, collecting retrospective data on specific cohorts 3-4 years after graduation. Using older waves of this survey, the existing literature reports substantial differentials related to family background and

² However, a national administrative data archive on university students' careers – not including information of family background – has been recently constructed by the Ministry of Education.

prior schooling (Di Pietro 2004; Di Pietro, Cutillo, 2008; Cingano, Cipollone 2007). Instead, Aina (2013) analyses the European Community Household Panel data with hazard modelling. Using administrative data from large universities, other studies analyse dropout and change of programs with competing risks (Clerici *et al.*, 2014) or latent class modelling (Belloc, 2012). All studies highlight the importance of prior schooling and the role played by socioeconomic background over and beyond prior schooling, and report different patterns by field of study and degree programs.

Cappellari, Lucifora, (2009) and D'Hombres, (2007) study the impact of the Bologna process and show that the reform has contributed to a small reduction of the dropout probability. Mealli, Rampichini (2012) analyse the effect of grants with regression discontinuity design, and show that at the threshold, grants contribute preventing dropout. Carrieri *et al.* (2015) analyse the influence of admission policies on retention and find that a stronger selection at entrance considerably reduces dropout risks. The literature on time to degree is scant. Aina *et al.* (2011) highlight the role of individual and family factors, and find that weak labour market prospects contribute to lengthening time to degree, while Garibaldi *et al.* (2012) observes a negative relation between tuition costs and timely completion probability.

Role of labour market conditions

According to theoretical predictions, bad economic conditions should increase education participation rates by reducing the opportunity costs of studying. Instead, the effect on the dropout probability is considered ambiguous. High unemployment rates may contribute to decreasing dropout rates if the opportunity costs of attending university decrease; however, the lack of income may cause financial difficulties and foster student dropout. Yet, poor labour market prospects may affect motivation and produce discouragement, contributing to reduce both enrolment and retention rates. In analysing the effects of the recent economic downturn in USA, Long (2013) and Hillman, Orians (2013) find counter-cyclical effects on university and community college enrolment. By exploiting territorial variability, Smith, Naylor (2001) found increasing dropout risk with higher unemployment in the UK. The few papers on Italy find inconclusive evidence on both enrolment and dropout. Di Pietro (2006) finds a negative relation between unemployment and dropout rates. Instead, Di Pietro (2004) reports a negative effect of local unemployment rates on enrolment and no effect on dropout, while in Di Pietro and Cutillo (2008) the effect is still insignificant on dropout but positive on enrolment.

3. The Italian educational system

In the Italian schooling system formal education starts at age 6 and is compulsory until age 16. Children attend eight years of comprehensive schooling, five years of primary education and three years of lower secondary education. Lower secondary school ends with a national examination at age 14, after which children choose between a numbers of upper secondary school programs, broadly classified into academic, technical and vocational tracks. The different educational programs differ substantially in curricular content and academic standards. Lyceums (making up the academic track), widely considered the natural path to university, are generally the most demanding. Technical institutes provide intermediate level education, with substantial academic content, but more labour market oriented. Vocational institutes, alongside general education, offer school-level training for low-level technical jobs. The high school diploma (*maturità*) is attained after 5 years and gives access to university.³ The Italian schooling system is mainly public: private institutions host only a small share of the student body. Private schools at the upper secondary level often have a remedial character, hosting affluent children who experienced school-year failures, and offering lower quality instruction than public schools (Bertola *et al.* 2008).

Despite the strongly stratified character of upper secondary education, in the Italian system there are no ability-related admission restrictions, neither at the transition between lower and upper secondary school – students can freely choose between lyceums and other school-types offering more vocationally oriented curricula – nor at the transition to higher education. All students possessing a 5-year high school diploma have unconditional access to university (although for some fields of study, enrolment is limited and regulated by admission tests).

In the Italian higher education system there is no formal divide between academic programs and polytechnic higher education, so we will refer to higher education and university system as synonymous. The system is mainly made by public universities, although there are a few prestigious private institutions. With these exceptions, reputation is not particularly important in Italy, because whatever the institution delivering the degree, it has the same “legal value”. Excluding the law and medical schools and few other degree-programs lasting 5-6 years, since the implementation of the EU ‘Bologna process’ in 2001, the system has been organized into a 3-year bachelor program followed by a 2-year master program. All students enrol in one

³ As an alternative to state-level education, students may also enter regional-based vocational training, lasting 2-3 years and not allowing access to higher education.

particular field (for example economics, physics or philosophy), and have limited leeway on which courses to follow. If students fail an exam or get a poor grade, they may try again until they obtain a result that they consider satisfactory. There are no official limits to time-to-degree, so students often take much longer than the official study length to complete their studies – in 2014, the average time to completion of 3-year programs was 4.6 years (Almalaurea 2014).

Tuition costs in public universities have limited variation across institutions and fields of study, although there are significant geographical differences.⁴ Similar to other European countries, they are generally low, and depend on per-capita household income; it is therefore unlikely that low-income students will not enrol because of direct costs of education. Financial aid in the form of grants is limited.⁵ However, subject to budget constraints, some scholarships are provided to low-income students obtaining a given amounts of credits and a given minimum grade point average.

4. Data and methods

We employ the Survey on High-School Graduates, carried out by the Italian National Statistical Institute (ISTAT) each 3-4 years on students completing secondary school, with the purpose to study educational and labour market experiences of Italian upper secondary degree holders. We use the latest available wave, carried out in 2011, interviewing graduates 2007. The survey provides information on educational and labour market histories, prior schooling and family background. Our interest rests on the behaviour of young individuals following a relatively regular schooling career, so we analyse individuals aged 21 or less at high school completion.

Given the long time span and the retrospective character of the data collection, the questionnaire focuses on first and current spells in education and work, not enabling to fully reconstruct the timing of the events of interest and unambiguously relate educational and working careers. Moreover, the question wording does not allow identifying precisely the dropout event or the exact moment of degree completion. In this light, we analyse the occurrence of the events of interest within the observed time span of 4 years, rather than the timing of the events.

⁴ Tuition costs in Northern public institutions are on average 17% higher than in Southern institutions (Federconsumatori, 2015).

⁵ In Italy 19% of the students benefitted of public financial aid in 2010-11, a significantly lower share than in most other European countries (OECD, 2013).

We analyse university enrolment together with two distinct educational outcomes: retention (vs. dropout) within 4 years from enrolment and, limited to those enrolled in 3-year programs, timely completion, defined as degree attainment within 4 years.⁶ Our analytical strategy can be described as follows. Let E be the binary variable indicating university enrolment, D the binary variable indicating dropout from the university system, and C the binary variable indicating timely degree completion for those enrolled in a 3-year program ($P_3=1$). In a first step, we model $P(E=1)$, $P(D=1|E=1)$ and $P(C=1|E=1, P_3=1)$ with separate logistic regression, with the aim to characterize the determinants of each event of interest, and the – direct and indirect – role played by social origin.

We adopt this descriptive approach because we believe the most relevant question is: *How do enrolled students from different family backgrounds with the same prior schooling history behave?*

Differently from other scholars, we do not attempt to estimate the effect of social origin *net* of all other observable and unobservable individual characteristics. For example, Montmarquette (2001), Di Pietro (2003) and Di Pietro, Cutillo (2008) model the joint likelihood of enrolment and dropout with bivariate probit, to account for the fact that enrolment and dropout decisions may be influenced by the same individual-specific unobserved factors. Similarly, Cingano, Cipollone (2007), use sample selection methods, acknowledging that the dropout probability is estimated on the subgroup of university entrants, so children of lower backgrounds are likely to be positively selected in terms of unobservable characteristics like innate ability or motivation.

In our view, the comparison between individuals of different family backgrounds with identical prior schooling history and identical unobserved personal trait is not particularly salient. The reason is that it is very unlikely that individuals of different family background with identical innate ability or motivation will experience identical (prior) educational careers (see the Appendix for a more detailed argumentation).

In a second step, we examine jointly the evidence on enrolment and dropout, and plot the estimates of the retention probability $P(D=0|E=1)$ against the enrolment probability $P(E=1)$, for subgroups of children defined by socio-demographic characteristics and features of prior schooling. This simple strategy is particularly useful, since it allows us to visualize the degree to which disadvantaged groups with respect to

⁶ For students entering university immediately after high school we observe dropout and degree completion occurring within 4 years from enrolment. We focus on these students, because for those delaying entry the time span is shorter and completion is rarely observed.

university enrolment are also disadvantaged with respect to persistence, and to get a sense of the joint probability of enrolling and not withdrawing within 4 years $P(E=1, D=0)$. To the extent that not all students entering higher education ultimately attain the degree, in the perspective of social selection this joint outcome is more salient than the enrolment probability per se. We will show that the disadvantaged groups in one dimension are also disadvantaged in the other, and that since the effects are cumulative, the differences across groups in the joint probability become impressive. Finally, focusing on students first enrolled in 3-year programs, we examine the relation between the two higher education outcomes retention and timely completion, by plotting the estimates of $P(D=0|E=1)$ and $P(C=1|E=1)$ for the same subgroups. This approach allows assessing whether social origin and the other socio-demographic characteristics affect retention and timely completion in a similar way, or if the two outcomes are driven by different mechanisms.

In a third step, we add the youth provincial unemployment rate to the models, to analyse whether and how labour market conditions relate to enrolment, dropout and timely completion, and whether labour market conditions contribute to explain statistically the observed territorial differentials.

5. Variables and descriptives.

5.1 Explanatory variables

To analyse the determinants of the educational outcomes of interest we include four sets of explanatory variables: sociodemographic characteristics, prior schooling, university studies and contextual variables. To study enrolment decisions, we first consider sociodemographic characteristics and then add prior schooling to assess the extent to which sociodemographic differentials are explained by prior schooling characteristics. Finally, we add contextual variables. To study dropout and timely completion we also include university variables.

Sociodemographic variables

We consider gender, macro-area of residence during secondary school (North, Centre and South) and family background, measured by parental education and parental class. We define parental education as the highest educational level of mother and father. Similarly, we define parental class according to the highest level of occupation. We use the restricted EGP (Erikson, Goldthorpe and Portocarero) classification that

distinguishes between service class (managers and professionals), intermediate, and working class (manual and unskilled workers).⁷

Prior schooling characteristics

We include the upper secondary school-type (lyceum, technical school, vocational school) and marks obtained at the lower and upper secondary school final examinations. We also consider features of the schooling career signalling fragility: whether the student has failed some subjects or repeated a school year, and whether she has changed institution during upper secondary school (usually students change school if they do not meet the school's performance standards; Contini and Triventi, 2015). Finally, we consider whether the secondary school institution at graduation was public or private.

University studies

We consider characteristics of the university and educational programs: field of study, degree length, macro-area of the university, and whether the student has changed region to attend university. In addition, we consider whether respondents were working while studying during the first academic year at university, and as an indicator of motivation, if they declared to be determined to enter university at the end of high school.

Contextual variables

On the demand side, we include the youth unemployment rate at provincial level to analyse the effect of labour market conditions on participation to tertiary education and educational outcomes. On the supply side, and limited to the enrolment decision model, we included a measure of diversity of the degree- programs in the area of residence in terms of number of different fields of study. This variable always turns insignificant, so we will omit further references to it.

5.2 Descriptive statistics

The share of high school graduates who enrolled at university within four years after graduation is 65.2%, the majority of which enrolled immediately after high school (nearly 88%). Among students enrolling immediately after graduation, those reporting withdrawal before degree completion are 12.3%. This share does not include transfers to other degree programs (approximately 10.5%).

⁷ Age is not included, because it is nearly collinear to school year repetitions. Due to the limited number of immigrant background students in the data, we also ignore the migrant status.

Table 1. Descriptive statistics on enrolment, dropout and timely degree completion

	Total
Enrolled (within 4 years) (%)	65.2
Of which: Enrolled right after graduation (%)	87.6
Dropout within 4 years among enrolled right after graduation (%)	12.3
Degree attained within 4 years among enrolled right after graduation in 3-year program (%)	29.8

Dropout rates are considerably higher in the first year (Table 2). Moreover, dropout rates for individuals delaying university entry are larger than for immediate entrants: the first-year probability increases from 8.6 for those enrolled in 2007 to 24.6 for those enrolled in 2009. A similar pattern is observed for the hazard rate.⁸

Table 2. Dropout probability and hazard rates, by year of university enrolment

	year of enrolment					
	2007		2008		2009	
	Probability	Hazard	Probability	Hazard	Probability	Hazard
Dropout 1° year	8.6	8.6	17.3	17.3	24.6	24.6
Dropout 2° year	2.7	3.0	4.0	4.8	-	-
Dropout 3° or 4° year	1.0	1.1	-	-	-	-

Descriptive statistics are shown in Table 3. In the first two columns we report the distributions of the explanatory variables among high school graduates and university entrants. These figures give information on the selectivity process between upper secondary and tertiary education. In the last three columns, we show the proportion of enrolled, dropouts and timely degree completions.

⁸ Dropout rates obtained by all waves of the Survey on High-School Graduates are lower than the corresponding aggregate rates reported by the Ministry of Education – for example, the official 1-year dropout rate was around 16% for 3-year programs and 9% for 5-year programs in 2008-9 (ANVUR, 2013). In analysing earlier waves of the survey, Cingano and Cipollone (2007) offer some potential explanations of the discrepancy, suggesting that there might be some misreporting in the survey data. In particular, some students might omit declaring university enrolment if it is a transient state not followed by actual participation, or they might omit declaring dropout if they hope resuming their studies in the near future. We find no empirical support for the former explanation, as enrolment rates in the survey are very similar to official ones. Instead, we agree with Cingano, Cipollone on the potential relevance of the latter, but we cannot check it empirically. We suggest two additional explanations: (i) Official dropout rates refer to the entire student-body, while our figures refer to the subsample of students enrolled within few years after high school graduation. There is wide evidence that older students and working-students, accounting for a significant share of the Italian university student population, have substantially higher dropout rates. (ii) The survey sample does not seem fully representative in terms of student ability and motivation, as the final high school examination grade distribution is more favourable among survey participants than according to official statistics. Overall, since disadvantaged students are less likely to enter the survey, we will consider our results as conservative estimates of socioeconomic inequalities.

Table 3. Descriptive statistics

	% among		% enrolled among HS graduates	% dropout among enrolled right after HS	% timely completion among enrolled in 3-year programs right after HS
	HS graduates	Enrolled			
Male	47	43	58.8	14.8	25.0
Female	53	57	70.9	10.5	33.5
Parental ed. Lower secondary	32	23	46.7	19.0	24.0
Parental ed. Higher secondary	49	52	69.6	12.4	29.7
Parental ed. Tertiary	19	25	84.8	6.6	35.7
Working class	29	22	47.8	17.3	23.5
Intermediate	46	46	65.6	13.2	30.0
Service class	24	32	85.6	8.0	33.9
Vocational track	16	6	23.6	36.7	15.0
Technical track	49	43	57.6	18.3	23.0
Lyceum track	35	51	94.0	5.6	37.3
HS final grade 60-69	32	22	44.9	22.6	14.7
HS final grade 70-79	27	26	62.8	13.8	22.4
HS final grade 80-89	19	22	75.4	10.7	30.6
HS final grade 90-100	22	30	87.8	5.8	45.9
D grade in MS	16	8	33.6	27.6	15.5
C grade in MS	31	26	53.0	19.4	18.6
B grade in MS	27	31	73.7	11.9	28.7
A grade in MS	26	35	90.4	5.3	41.9
Change in secondary school	7	6	53.5	31.0	17.8
No change	93	94	66.0	11.4	30.5
Private HS	5	6	72.3	15.5	28.3
Public HS	95	94	64.8	12.2	29.9
Repetition or subject failure	50	43	55.2	18.5	19.1
Regular career	50	57	75.3	8.1	37.4
Intended to enrol	64	91	92.8	9.9	30.9
Not intended to enrol	36	9	15.6	54.2	11.9
Area of origin North	37	37	65.2	11.2	42.6
Area of origin Centre	19	19	66.1	12.6	29.6
Area of origin South	44	44	64.7	13.2	18.5
5-year program		19		8.6	-
3-year program		81		13.2	29.8
Working student		7		24.2	19.0
Full-time student		93		11.5	30.6
Changed region		18		8.4	38.1
No change		82		13.2	28.0
Area of University North		40		10.8	42.2
Area of University Centre		22		11.2	29.4
Area of University South		37		14.7	15.9

6. Role of socio-demographic characteristics and prior schooling.

Since our aim is to investigate social selectivity, we first estimate a model including only socio-demographic characteristics: the variable of major interest is family background, measured by parental education and class (Model I in Table 4). We then add variables on prior schooling: as highlighted above, they are in themselves outcomes of a previous strong socially selective process (Model II). Finally, in the dropout and timely completion models we include university variables (Model III). Results in terms of Average Marginal Effects (AME) are summarized in Table 4.⁹

Gender

Gender differences in the enrolment and dropout probabilities vary across macro-areas. Females are more likely to enter university everywhere, but this advantage is particularly marked in Southern Italy: given socioeconomic background, in the South females have an advantage over males of almost 16 percentage points. The gap is almost entirely explained by the more proficient schooling career of girls in the North, whereas in the rest of the country girls remain substantially more likely to enrol, even after controlling for prior schooling. Females have lower dropout chances than males: this gap is larger in the North (6 percentage points), than in the Centre-South (3-4 percentage points). The gender differential is fully explained by prior schooling and university variables in the Centre-South, but not in the North. As for timely completion, girls are more likely to attain the degree within 4 years than males (nearly +10 percentage points), but this gap is fully explained by prior schooling and the characteristics of university and degree-program.

Geographical area

Geographical differences in enrolment rates are small for females and substantial for males; given prior schooling, in the Centre and South males exhibit a lower probability to enter university than their Northern peers (4-5 percentage points). Dropout probabilities are larger among females from the Centre-South than among those from the North, but vary little for males; however, the chances of dropout given all controls become substantially larger for students from the Centre. Impressive geographical differences are observed

⁹ The AME corresponding to a continuous independent variable is the average over all sample units of the first derivative of the estimated density function with respect to that variable. If the explanatory variable is categorical, AME measures the discrete change relative to a comparison with the reference category. Differently from logit and probit regression coefficients, AME are comparable across nested models (Mood, 2010; Wooldridge, 2012).

for timely degree attainment, which is much more likely in the North for both genders: given prior schooling and university variables, the average probability difference between North and South is 18 percentage points.

Social origin

Social origin inequalities are substantial on all outcomes. Parental education and class play similar roles on enrolment: on average, the advantage of the highest strata over the lowest one is 25 percentage points for each dimension. Hence, the gap reaches 50 percentage points when comparing a youngster of highly educated parents in the service class and one with lowly educated parents in the working class. The advantage decreases but remains substantial when comparing individuals with the same prior schooling history: on average, the advantage of the highest strata over the lowest ones is nearly 8 percentage points on parental education and nearly 12 percentage points on social class. Social origin also strongly affects the dropout probability, but parental education plays a stronger role than parental class. This pattern does not change qualitatively when adding university variables. Parental education has also a large effect on the timely completion probability (+ 4 percentage points between the highest and lowest strata), while parental class plays virtually no effect.

Prior schooling

Prior schooling characteristics are the strongest predictors of university enrolment and success. The upper secondary school-type is the major divide: students with a *lyceum* diploma have considerably higher enrolment, retention and timely completion probabilities than those with a technical diploma, and the difference is even larger if we compare them to vocational schools graduates. On average, the enrolment probability of a *lyceum* graduate is 55 percentage points higher than that of a student from the vocational track and 29 points higher than a student from the technical track. Corresponding differences in the dropout probability are 25 and 10 percentage points, in the timely completion probability 23 and 12. Proficiency in both lower and upper secondary school are also highly relevant. Finally, *ceteris paribus*, students previously attending private schools or having experienced school failures (repetition or debts) do not differ much from other students in terms of enrolment behaviour, but have much higher chances of withdrawing (+ 3-7 percentage points) and lower chances of timely degree attainment (-3-6 percentage points).

Table 4. Models for enrolment, dropout and timely completion probabilities. Individual factors. (Average marginal effects).

Models	ENROLMENT		DROPOUT			TIMELY COMPLETION		
	I	II	I	II	III	I	II	III
<i>GENDER & AREA</i>								
Area of origin for Males (ref. North)								
Centre	-0.046***	-0.054***	0.011	0.025**	0.054**	-0.131***	-0.162***	-0.121***
South	-0.022**	-0.041***	-0.001	0.021**	-0.009	-0.219***	-0.263***	-0.188***
Area of origin for Females (ref. North)								
Centre	0.001	-0.028***	0.024**	0.038***	0.066***	-0.139***	-0.156***	-0.102***
South	0.055***	-0.003	0.025***	0.047***	0.019	-0.252***	-0.284***	-0.180***
Female for North (vs. Male)	0.086***	0.015*	-0.067***	-0.038***	-0.030***	0.112***	0.053***	-0.005
Female for Centre (vs. Male)	0.133***	0.041***	-0.054***	-0.026**	-0.018	0.104***	0.059***	0.015
Female for South (vs. Male)	0.163***	0.052***	-0.040***	-0.012	-0.002	0.080***	0.032***	0.003
<i>SOCIAL BACKGROUND</i>								
Parental education (ref. tertiary)								
lower secondary	-0.244***	-0.077***	0.107***	0.063***	0.059***	-0.090***	-0.043***	-0.045***
higher secondary	-0.087***	-0.013	0.050***	0.028***	0.024***	-0.054***	-0.023**	-0.026**
Parental class (ref. service class)								
Working	-0.255***	-0.116***	0.038***	0.020**	0.021**	-0.043***	-0.007	-0.009
Intermediate	-0.143***	-0.069***	0.020***	0.011	0.010	-0.007	0.012	0.018*
<i>PRIOR SCHOOLING VARIABLES</i>								
Secondary school (ref. <i>lyceum</i>)								
Vocational		-0.553***		0.250***	0.251***		-0.196***	-0.212***
Technical		-0.291***		0.105***	0.103***		-0.120***	-0.122***
Mark high school (10 points)		0.087***		-0.035***	-0.035***		0.085***	0.092***
Mark middle school (ref. D)								
A		0.091***		-0.036***	-0.035***		0.064***	0.063***
B		0.062***		-0.013	-0.013***		0.024	0.021
C		0.039***		-0.006	-0.006		-0.009	-0.013
Changed school		0.007		0.067***	0.069***		-0.026	-0.035*
Private school		-0.012		0.060***	0.064***		-0.040**	-0.060***
Repetition or debt		-0.012**		0.034***	0.036***		-0.034***	-0.040***
<i>UNIVERSITY VARIABLES</i>								
5-year vs. 3-year program					-0.022**			-
University in different region					-0.016**			0.040***
Area of University (ref. <i>North</i>)								
Centre					-0.021			-0.044**
South					0.037*			-0.107***
Working student					0.068***			-0.087***
Field of study					Yes			Yes
N° Observations	25,042	25,042	12,038	12,038	12,038	10,048	10,048	10,048
Pseudo R-squared	0.0909	0.2956	0.0340	0.1530	0.1688	0.0619	0.1654	0.2048

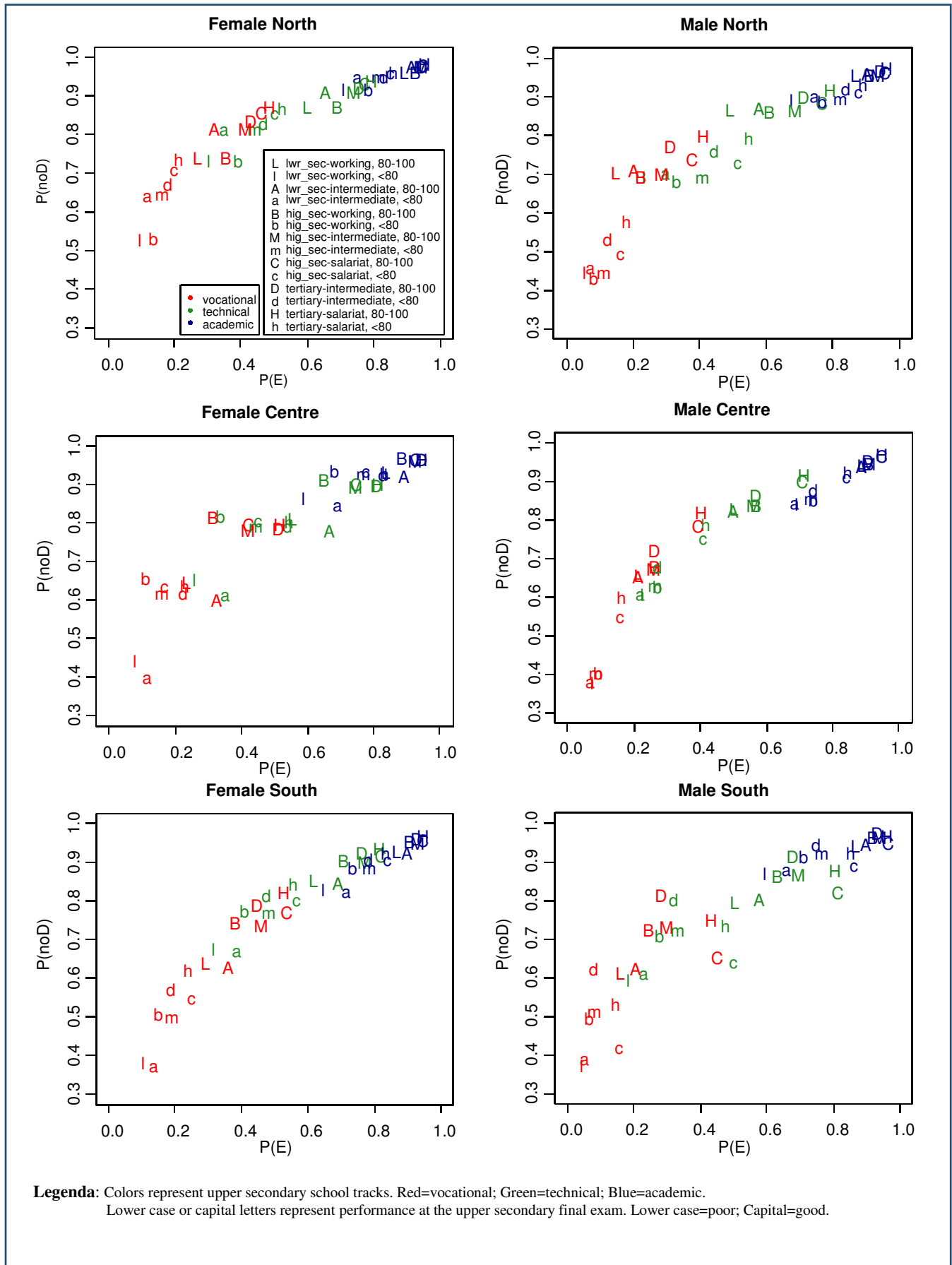
*p-value<0.05; ** p-value<0.01; *** p-value<0.005

7. An overall picture of social selectivity

In order to visualize the cumulative effect of social background and prior schooling career, we put together the evidence on enrolment and retention probabilities in a graphical representation (Figure 1). We plot the estimates of the retention (no dropout) probability $P(D=0|E=1)$ against the enrolment probability $P(E=1)$, for subgroups of children defined by parental background (parental education and class), upper secondary school track, upper secondary final grade (60-79, 80-100), geographical area of the upper secondary school and gender. For the sake of clarity, we plot separate graphs by the latter two characteristics. Figure 1 conveys two main messages:

- 1) There is a very strong correlation between the enrolment probability and the retention probability across categories of students based on their observed characteristics. The subgroups of students more likely to enrol are also more likely not to withdraw from university. Since the effects are cumulative, differences across groups become dramatic.
- 2) Differences are substantial even when we compare children with similar prior schooling. Children of low social origin are disadvantaged in both enrolment and retention, in particular among those who attended the vocational or technical tracks. Somewhat lower differences hold for children with an academic upper secondary degree. If we acknowledge that high school choices are strongly dependent on social origin, and that those who do not attain a high school degree (approximately 20% in 2007) disproportionately come from the lower classes, social selectivity appears even more severe.

Figure 1. Estimated enrolment probability (x-axis) against retention probability among enrolled (y-axis), by gender, geographical area, upper secondary school characteristics and performance.



Letters stand for parental background: L=low parental education and occupation; M=intermediate parental education and occupation. H=high parental education and occupation. A, B, C D are mixed cases (e.g. low parental education and intermediate parental occupation).

More specifically, our findings can be sketched as follows:

- The enrolment probabilities spread along the entire 0-1 interval, and the retention probabilities vary between less than 0.4 to nearly one. Thus, the joint probability to enter university *and* not withdraw within the observed time-span varies enormously across categories: between 0.02 to nearly one.
- Differences according to prior schooling, as expected, are very large: students with a vocational diploma (red) have the least positive outcomes, and those from lyceums (blue) the most favourable ones. The same occurs for students with low grades (lower case letter), who have less favourable outcomes on both enrolment and dropout as compared to those with high grades (capital letter), from all school-types.
- Social background differences given socio-demographic characteristics and prior schooling are remarkable. Take as an example Southern females, with high grades and vocational diploma: for the lowest stratum (*L*) the enrolment probability is 0.3 and the retention probability is 0.64, hence they have a 0.19 probability of enrolling and not withdrawing. For the highest stratum (*H*) the enrolment probability is 0.55 and the retention probability 0.83, i.e. a 0.46 joint probability. Take instead the low performing from lyceums: among those from the lowest stratum (*l*) the enrolment probability is 0.63 and the retention probability 0.82, i.e. a 0.52 joint probability, while for the highest stratum (*h*) the enrolment probability is 0.85 and the retention probability 0.95, i.e. a 0.81 joint probability. Differences are less marked, but still substantial, among the well performing students from lyceums: for the lowest stratum (*L*) the enrolment probability is 0.9 and the retention probability 0.93, i.e. a 0.84 joint probability, while for the highest stratum (*H*) the enrolment probability is 0.95 and the retention probability 0.97, yielding to a 0.92 probability of enrolling and not withdrawing.¹⁰

¹⁰ Probability differences between subgroups in these graphs differ from what we can deduce by inspecting the AME in Table 4. The reason is that the underlying models are not identical: here they are by gender and geographical area, but include fewer explanatory variables. Hence, while in step 1 we compared individuals with the ‘same schooling history’, we refer here to children with ‘similar prior schooling’.

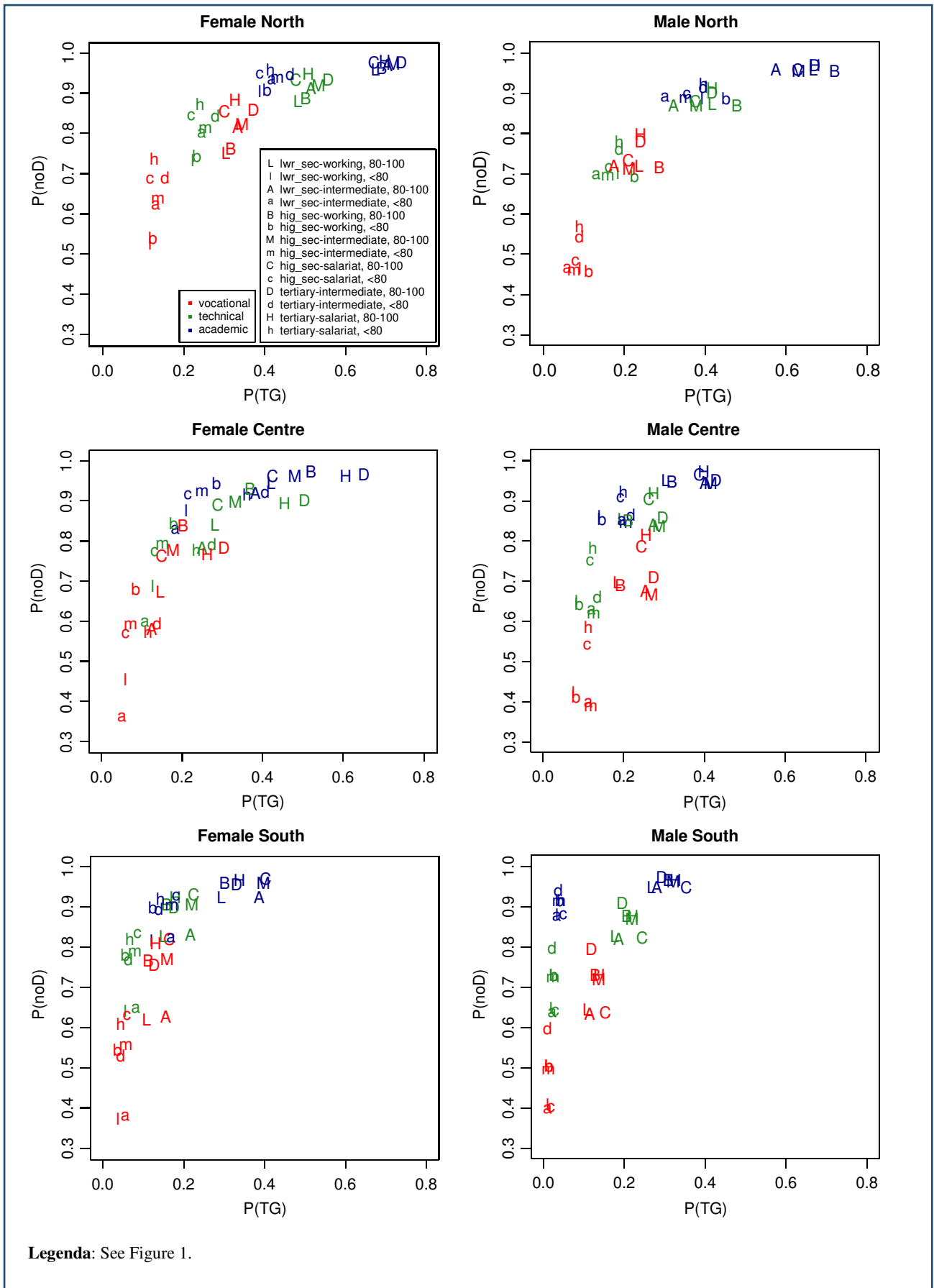
8. Retention and timely completion: two distinct outcomes

Limiting the focus to students enrolled in 3-year programs, for whom we observe degree attainment within 4 years, we now examine the relation between the two alternative outcomes of higher education under study: the retention probability $P(D=0|E=1)$ and the timely completion probability $P(C=1|E=1)$. By inspecting Figure 2, we see that the relation between retention and timely graduation is positive, but curvilinear. In most sub-graphs, there are well-defined clusters by school-type and grades. Indeed, subgroups performing better in one dimension tend to do better also in the other one. However, we find much larger social background differences given prior schooling along the retention probability axis than along the timely completion one (with the exception of males in the North), in particular for those with weak educational backgrounds.

Our speculative explanation of this body of evidence is that timely completion is a much more stringent outcome than retention. While the first implies that students passed the exams and accumulated credits, the second only states that students are still enrolled. Yet, retention may coexist with very little progress in credit earning. In this light, the low probability of timely completion and its limited variability for students coming from vocational secondary schools or from technical schools with low grades, points to an actual difficulty experienced by students with poor academic past credentials. On the contrary, we find much larger variability across subgroups of children in the retention probability across social backgrounds, because dropout is a much more severe outcome for families of high socioeconomic background than it is for those of low social origin. To state it explicitly, to avoid quitting you just need to pay tuition fees, while to attain the degree - and do it quickly - you need a lot of work.

Focusing on territorial differentials, we observe huge differences on timely completion, but not on retention. Timely completion rates are almost nil for all males in the South with low grades, whatever the school type, and somewhat higher but still very low for southern females; instead, they are relatively high in the North, for the students with a good high school background.

Figure 2. Estimated timely completion probability (x-axis) against retention probability (y-axis) among enrolled in 3-year programs, by gender, geographical area, upper secondary school characteristics and performance.



9. Role of labour market conditions.

Previous analyses show that higher education outcomes markedly differ along the North-South divide, even after controlling for sociodemographic features, prior schooling and university characteristics. Therefore, context factors matter. As highlighted above, labour market conditions vary considerably across the country, the South being disadvantaged with respect to unemployment rates, female labour market participation and poverty rates. In an exploratory perspective, we now analyse whether and how labour market conditions relate to enrolment, dropout and timely completion, and whether labour market conditions contribute to explain statistically the observed territorial differentials. To measure labour market conditions we use youth unemployment rates at the provincial level in 2007.¹¹ This variable is added to the richer models in Table 4 (model II for the enrolment equation and to model III for the dropout and timely completion equations). We include unemployment rates in quadratic term to capture possible non-linear effects. Our main results, shown in Table 5, can be summarised as follows:

- (i) Unemployment rates follow an inversely U-shaped relation with enrolment, a negative relation with timely completion, and a positive relation with dropout probabilities.

While at low unemployment rates individuals increase university participation as unemployment rates increase, we observe the opposite pattern when unemployment rates become larger.¹² Hence, with poor labour market conditions discouragement seems to prevail. Similarly, in high unemployment areas individuals are more likely to dropout from university. This result highlights that in Italy, even before the economic downturn of 2008-2009, university dropout was not due to labour market acting as a *pull* out factor. In this light, although positive accounts of withdrawal are possible in principle, they are unlikely to hold in the context under study. Instead, the timely completion likelihood decreases as unemployment rises: contrary to our results on enrolment and dropout, this is consistent with the rational behaviour view that individuals should engage more in education with poor labour market conditions.¹³

¹¹ Youth provincial unemployment rate (age15-29), released by the National Statistical Institute (110 provinces). In 2007, this rate varied across provinces between 3% and 46%. The average provincial rate was 12.2% in the North, 16.0% in the Centre and 31.3% in the South.

¹² Given the large territorial differences in unemployment rates, (cf. note 11), when estimating separate models by macro-area we find a positive effect of unemployment rates in the North and a negative effect in the Centre and South.

¹³ We also estimated models for dropout and timely completion including the unemployment rate of the university province instead of the province of origin. This rate has a slightly higher explanatory power for dropout and a slightly lower explanatory power for timely completion, but results are very similar.

- (ii) Unemployment rates partially explain the territorial differentials in the timely completion probability, fully explain the male enrolment gap between North and South, but do not contribute explaining territorial differences in dropout rates.

More specifically, territorial AME for timely completion are very large, but once we include the unemployment rate in the model, the observed gaps decrease by 25-45%. Unemployment rates fully explain the enrolment North-South differential for males, and partially explain the enrolment North-Centre differential for both genders, but does not reduce the large dropout North-Centre gap. According to our estimates, if exposed to similar unemployment rates, the enrolment probability of Southern females would exceed that of Northern females and the dropout probability of Northern males would exceed that of Southern males.

Table 5. Models for enrolment, dropout and timely completion. Role of labour market conditions.

	ENROLMENT	DROPOUT	TIMELY COMPLETION
PANEL 1			
Logit regression coefficients			
Unemployment Rate	0.0162**		-0.0242***
Unemployment Rate ²	-0.0006***	0.0004***	
SHAPE	Inverse U shaped (decreasing from rate 13%)	Increasing	Decreasing
Pseudo R-squared	0.2966	0.1707	0.2078
PANEL 2			
AME North-South (from Table 4)			
Males	0.041***	0.009	0.188***
Females	0.003	-0.019	0.180***
<i>Controlling for unemployment rate</i>			
Males	0.008	0.042**	0.112***
Females	-0.029***	0.010	0.103***
AME North-Centre (from Table 4)			
Males	0.054***	-0.054***	0.121***
Females	0.028***	-0.066***	0.102***
<i>Controlling for unemployment rate</i>			
Males	0.048***	-0.045*	0.089***
Females	0.023***	-0.061***	0.070***

NOTES. 1. All models control for sociodemographic characteristics and prior schooling. Dropout and timely completion models also include university variables. 2. Unemployment rate: youth unemployment rate by province.

*p-value<0.05; ** p-value<0.01; *** p-value<0.005

10. Summary and conclusions

In this paper, we employ the latest wave of the Survey on High-School Graduates, to analyse university enrolment and two related outcomes: retention vs. dropout within 4 years from enrolment, and timely completion. We model the occurrence of each of these events in order to improve the understanding of the role of social origin, through and on top of inequalities established in previous stages of the schooling career. Prior schooling and performance stand out as strong predictors of all higher educational outcomes; yet, social background appears to be a relevant factor even after controlling for these characteristics.

Secondly, we jointly analyse the evidence delivered by these estimates, by plotting the estimated retention probability against the enrolment probability, for subgroups of individuals identified by socio-demographic characteristics and prior schooling. This simple strategy allows visualising the degree to which disadvantaged groups with respect to enrolment are also disadvantaged with respect to persistence. We find a strong positive relation between enrolment and retention probabilities: the overall picture highlights impressive social origin inequalities, even when comparing individuals with similar previous school background.

We then analyse the relation between retention and timely completion. Subgroups performing better in one dimension tend to do better also in the other one; however, we find much larger social background differences (given prior schooling) in retention. Our speculative explanation is that in the Italian system timely completion is a much more stringent outcome than retention. In this light, the low probability of timely completion of students who attended technical schools with low grades or vocational education, and its limited variability across social backgrounds, point to an actual difficulty experienced by students with poor academic past credentials. Instead, we find more variability across social backgrounds in the retention probability: this suggests that among families of high social origin dropout is a very undesirable outcome, to avoid as far as possible. Whether students from high social background and poor school credentials ultimately drop out or just take more time to graduate cannot be assessed with these data.

From a policy perspective, the low chances of enrolment and successful completion of the students following educational programs with little academic content or with poor school records are clearly detrimental to equity (since these students come disproportionately from disadvantaged backgrounds), but could be

considered beneficial to efficiency. Yet, the degree acquisition of endowed students from disadvantaged backgrounds should be highly encouraged from both perspectives; it would foster equity, but it would also raise the share of individuals with higher education without lowering learning standards.

Thirdly, we analyse how labour market conditions relate to higher education outcomes and whether they contribute explaining the large geographical differentials. Enrolment, retention and timely completion probabilities are lower in provinces with high youth unemployment rates. These results altogether suggest that with poor labour market prospects individuals lose motivation and reduce their engagement in education. However, unemployment rates only contribute explaining geographical differentials in timely completion.

Why do individuals drop out of higher education in Italy at the onset of the economic crisis of 2008? This is a complex issue, difficult to address with general surveys data. However, our analyses provide some indications on the role of alternative factors potentially at play (Jordan *et al.* 1996; Watt, Roessingh 1994). Indeed, poor prior academic preparation stems out as major determinant of withdrawal. Instead, our results suggest that the labour market does not act as a major *pull* force diverting students from completing studies, since with better employment prospects there is less dropout.¹⁴ Moreover, as highlighted above, there is indirect evidence that high unemployment fosters discouragement, so low motivation may also be an important factor affecting students' educational careers. There is no evidence, instead, on the importance of alternative *pull* forces like family care duties, as the dropout probability is much larger for males than for females, whereas in Italy family care is still largely in charge of women.

¹⁴ Additional support to this interpretation is delivered by direct information on individuals' labour market experiences recorded in the survey: among dropouts who were not employed at time of enrolment, less than 20% started working within 6 months from withdrawal.

References

- Aina C., E. Baici, G. Casalone (2011) Time to degree: students' abilities, university characteristics or something else? Evidence from Italy. *Education Economics*, 19, 3, 311-325
- Aina C. (2013) Parental background and university dropout in Italy, *Higher Education*, 65, 4, 437-456
- AlmaLaurea (2015), XVII indagine Profilo dei Laureati 2014, Consorzio AlmaLaurea, Bologna, <http://www.almaLaurea.it/universita/profilo>
- ANVUR (2014) Rapporto sullo stato del sistema universitario e della ricerca 2013, Agenzia Nazionale di Valutazione del Sistema Universitario e della Ricerca, Roma http://www.anvur.org/attachments/article/644/Rapporto%20ANVUR%202013_UNIVERSITA%20e%20RICERCA_integrale.pdf.
- Arulampalam W., R. A. Naylor, J. P. Smith (2005) Effects of in-class variation and student rank on the probability of withdrawal: cross-section and time-series analysis for UK university students, *Economics of Education Review* 24, 251–262
- Belloc F., A. Maruotti, L. Petrella (2011) How individual characteristics affect university students drop-out: a semiparametric mixed-effects model for an Italian case study, *Journal of Applied Statistics*, 38:10, 2225-2239
- Bertola G., D. Checchi, V. Oppedisano (2008) Private school quality in Italy, *Giornale degli Economisti e Annali di Economia*, 66-3, 375-400
- Blundell R., L. Dearden, B. Sianesi (2001) Estimating the Returns to Education: Models, Methods and Results, Centre for the Economics of Education London School of Economics.
- Bourdieu P., J.C. Passeron (1990) *Reproduction in education, society, and culture*, Vol 4, Sage
- Boudon R. (1974) *Education, Opportunity, and Social Inequality: Changing Prospects in Western Society*, John Wiley & Sons
- Bound J., M. Lovenheim, S. Turner (2010) Why Have College Completion Rates Declined? An Analysis of Changing Student Preparation and Collegiate Resources, *American Economic Journal: Applied Economics*, 2(3), 129-57.
- Bowen W., M. Chingos, M. McPherson (2009) *Crossing the Finish Line: Completing College at America's Public Universities*, Princeton University Press, Princeton
- Bratti M., C. Broccolini, S. Staffolani (2010) Higher Education Reform, Student Time Allocation and Academic Performance in Italy: Evidence from a Faculty of Economics, *Rivista Italiana degli Economisti*, 2, 275-304
- Breen R. and J. Golthorpe, 1997 Explaining educational differentials: towards a formal rational action theory, *Rationality and Society*, 9, 275-305
- Cappellari L., C. Lucifora (2009) The “Bologna Process” and college enrollment decisions, *Labour Economics* 16, 638–647
- Carneiro, P., J. Heckman. (2002) The evidence on credit constraints in postsecondary schooling. *Economic Journal*. 112:705–734.

- Carrieri V., M. D'Amato, R. Zotti (2015) On the causal effects of selective admission policies on students' performances: evidence from a quasi-experiment in a large Italian university. *Oxford Economic Papers*
- Checchi D., C. V. Fiorio, M. Leonardi (2014) Parent's risk aversion and children's educational attainment, *Labour Economics*
- Cingano F., P. Cipollone (2007) University drop-out: The case of Italy, Temi di discussione Servizio Studi della Banca d'Italia, 626
- Ciriaci D. (2014) Does University Quality Influence the Interregional Mobility of Students and Graduates? The Case of Italy, *Regional Studies*, 48, 10, 1592-1608
- Clerici R., A. Giraldo, S. Meggiolaro (2014) The determinants of academic outcomes in a competing risks approach: evidence from Italy, *Studies in Higher Education*
- D'Hombres, (2007) The impact of university reforms on dropout rates and students' status: evidence from Italy, Joint Research Centre Scientific and Technical Reports, European Commission
- Di Pietro G. (2004) The determinants of university dropout in Italy: a bivariate probability model with sample selection *Applied Economics Letters*, 11:3, 187-191
- Di Pietro G. (2006) Regional labour market conditions and university dropout rates: Evidence from Italy, *Regional Studies*, 40:6, 617-630
- Di Pietro G., A. Cutillo (2008) Degree flexibility and university dropout: the Italian experience, *Economics of Education Review* 27, 546-555
- Dynarski S. (2003) Does aid matter? Measuring the effect of student aid on college attendance and completion, *American Economic Review*, 93(1), 279-288.
- Federconsumatori (2015) 6° Rapporto Nazionale Federconsumatori sui costi degli atenei italiani, [http://www.federconsumatori.it/news/foto/Rapporto%20Tasse%20universitarie%20Rapporto%202015-2016\(1\).pdf](http://www.federconsumatori.it/news/foto/Rapporto%20Tasse%20universitarie%20Rapporto%202015-2016(1).pdf)
- Garibaldi P. F. Giavazzi, A. Ichino, E. Rettore (2012) College cost and time to complete a degree: evidence from tuition discontinuities, *The Review of Economics and Statistics*, 94(3): 699-711
- Hillman N. W., E. Lee Orians (2013) Financial Aid's Role in Meeting State College Completion Goals. *Education Finance and Policy* 8:3, 349-363
- Ishitani T. T. (2006) Studying Attrition and Degree Completion Behavior among First-Generation College Students in the United States, *The Journal of Higher Education*, 77, 5, 861-885
- Jackson M. eds. (2013), *Determined to Succeed? Performance, Choice and Education*, Stanford University Press
- Jackson M., J. O. Jonsson (2013) Why does inequality of educational opportunity vary across countries? Primary and secondary effects in comparative context, in Jackson M. eds. *Determined to Succeed? Performance, Choice and Education*, Stanford University Press
- Jordan W. J., J. Lara, J. M. Mc Partland (1996) Exploring the Causes of Early Dropout among Race-Ethnic and Gender Groups, *Youth and Society*, 28,1 62-94.
- Kurlaender M., J. Jackson, J. S. Howell, E. Grodsky (2014) College course scarcity and time to degree, *Economics of Education Review* 41, 24-39.

Mealli F., C. Rampichini (2012) Evaluating the effects of university grants by using regression discontinuity designs, *Journal of the Royal Statistical Society A*, 175, 3, 775–798

Mood C. (2010) Logistic Regression: Why We Cannot Do What We Think We Can Do, and What We Can Do About It, *European Sociological Review*, 26, 1, 67–8

Montmarquette C., Mahseredjian S., Houle R. (2001). The determinants of university dropouts: a bivariate probability model with sample selection, *Economics of Education Review* 20 (2001) 475–484

OECD (2013), *Education at a Glance 2013: OECD Indicators*, OECD Publishing, <http://dx.doi.org/10.1787/eag-2013-en>.

Pigini C., S. Staffolani (2015) Beyond participation: do the cost and quality of higher education shape the enrollment composition? The case of Italy, *Higher Education* DOI 10.1007/s10734-015-9892-8

Watt D., H. Roessingh (1994) Some you win, most you lose: Tracking ESL dropout in high school (1988-1993). *English Quarterly*, 26, 5-7.

Smith J. P., R. A. Naylor (2001) Dropping out of university: a statistical analysis of the probability of withdrawal for UK university students, *Journal of the Royal Statistical Society A*, 164, 2, 389-405

Stinebrickner R., T. Stinebrickner (2008) The Effect of Credit Constraints on the College Drop-Out Decision: A Direct Approach Using a New Panel Study, *American Economic Review* 98,5, 2163-84

Stinebrickner R., T. Stinebrickner (2009) Learning About Academic Ability and the College Drop-Out Decision, NBER Working Paper 14810

Stinebrickner R., T. Stinebrickner (2013) Academic Performance and College Dropout: Using Longitudinal Expectations Data to Estimate a Learning Model, NBER Working Paper 1894

Tinto V. (1975) Dropout from higher education: a theoretical synthesis of recent research, *Review of Educational Research*, 45, 1, 89-125

Vignoles A. F., N. Powdthavee (2009) The Socioeconomic Gap in University Dropouts, *The B.E. Journal of Economic Analysis & Policy*: Vol. 9: Iss. 1 (Topics), Article 19.

Appendix.

Modelling dropout or completion with or without sample selection methods.

Consider the typical sample selection model (for simplicity, consider a continuous y):

$$z = \gamma + \delta w + u$$

(1a)

$$y = \alpha + \beta x + \varepsilon$$

(1b)

Let y be observed only if $z > 0$. (u, ε) are possibly correlated errors, so we may write $u = \theta + v_1$, $\varepsilon = \theta + v_2$, where v_1 and v_2 are mutually independent random disturbances, and independent of explanatory variables x and w . The parameter of interest is β , the causal effect of x on y , on the entire population. Various estimation methods are proposed (e.g. Heckman's two stage estimator) to account for the fact that the estimation of (1b) on the observed sample produces biased estimates of β .

In our case study, z is university enrolment (propensity) and y is dropout (propensity). x is a vector of explanatory variables including social background (x_1) and prior schooling characteristics (x_2). These variables are likely to affect enrolment as well, so $w \cong x$. θ can be conceived as innate ability or motivation. The parameter of interest in sample selection models is $\beta = E(y|x_1 + 1, x_2, \theta) - E(y|x_1, x_2, \theta)$. Yet, estimation of (1b) on the observed sample of university entrants provides information on:

$$E(y|x_1 + 1, x_2, z > 0) - E(y|x_1, x_2, z > 0) = \beta + (E(\theta|x_1 + 1, x_2, z > 0) - E(\theta|x_1, x_2, z > 0))$$

(2)

Due to sample selection, the second term is generally negative, because individuals from low social backgrounds are more positively selected in terms of unobserved traits. This expression provides an answer to the question: *How do university entrants from different family backgrounds with the same prior schooling history behave?* In our perspective, this is the quantity of main interest. Consider two ideal-typical young individuals just enrolled in university, one of low and one of high social background, having the same schooling history in terms of school-types, curricula and grades. The first has needed more effort than the

second one to obtain these results, so she will usually have better unobserved personal traits. Being aware of this difference, we are interested in comparing *their* dropout probabilities.

Instead, sample selection methods aim at comparing individuals sharing the same prior schooling history *and* unobserved characteristics, and answer the question: *How would individuals from different family backgrounds with the same prior schooling history, innate ability and motivation behave, if they enrolled in university?* While this question could provide useful information for prospective students wishing to evaluate their probability of success in university (and indeed it does inform on how success probabilities differ across social backgrounds), in our view it is not particularly salient to measure actual educational inequalities. We now attempt to explain why.

The general idea of causal reasoning is to produce counterfactual evidence, i.e. to compare identical individuals differing only in that some are exposed to a treatment and others are not. However, it is difficult to make a counterfactual argument on social origin. If we conceive parental education or social class as a “treatment”, we should acknowledge that exposition begins at birth. Two new born of different family backgrounds may be identical in terms of innate abilities. Yet, because of the “treatment”, as they grow older they start differentiating. Students of different family backgrounds may have identical schooling histories if low class individuals compensate their disadvantage with more effort or higher innate ability. Instead, individuals of different family backgrounds with identical unobserved personal traits may experience identical educational careers only by pure luck (i.e., due to the effect of the idiosyncratic random term component). Consequently, the contrast between individuals of different family backgrounds with identical prior schooling history and unobserved personal traits seems to be a rather odd one.¹⁵

There are two additional reasons for preferring a descriptive approach in this context. The first is that applying the sample selection model on high school graduates we are ignoring a previous stage of selection, because only a subgroup of the initial birth cohort eventually attains the high school diploma (approximately 80% in Italy). The second is that, as emphasized by Brandt and Schneider (2007) and Kennedy (2003; pg. 291), sample selection estimators perform very poorly if model assumptions are violated – the neglect of the first stage of selection is an example – or when the degree of collinearity between the explanatory variables

¹⁵ It is worth noticing that this argument may not apply to family income, as income is more variable over time and may be exogenously increased by scholarships or loans.

in the regression and the selection equations is high.¹⁶ This concern applies to our case study, since it is difficult to think of determinants of enrolment not influencing also the dropout and completion probability.

Additional References

Brandt P.T., C. J. Schneider (2007) So the reviewer told you to use a selection model? Selection models and the study of international relations, http://pages.ucsd.edu/~cjschneider/working_papers/pdf/Selection-W041.pdf

Kennedy P. (2003) *A Guide to Econometrics*, MIT Press.

¹⁶ We have also developed a simulation study confirming these findings. Results are available from the authors upon request.