# Do Financial Conditions Play a Role in University Dropout? New Evidence from Administrative Data

**Dalit Contini and Roberto Zotti**

**Abstract**  A large strand of research in the economics and sociology of education has highlighted the existence of deeply rooted inequalities in educational choices along socioeconomic lines, even when net of prior performance. These disparities may take different forms at different stages of schooling and across institutional systems. Yet, due to the lack of data, it is often difficult to disentangle the role played by the various dimensions of socioeconomic background on students' educational careers. While parental education and occupation may shape aspirations (and thus the wish to undertake ambitious educational programmes), lack of income could represent a material obstacle to the continuation of study. In this chapter, we focus on the effect of financial conditions on the probability of dropping out from university. Italy is an interesting study case, because the education system is mainly public and university tuition fees are relatively low and income progressive. Because direct costs for disadvantaged students are low, we would expect income not to be highly relevant in this context. By exploiting a unique data set from the University of Torino (in northern Italy) linking administrative data from students' university careers and information on parental characteristics collected at matriculation, we analyse how socioeconomic background influences the first-year dropout probability. While extremely relevant in earlier educational outcomes, parental education and occupation no longer exert a sizable effect at this point in students' lives. Instead, we find that economic conditions greatly influence the chances of completing university. This result suggests that low tuition fees may be insufficient to foster the participation of low-income high school graduates and that additional forms of support might be needed to ensure equity and, at the same time, raise the share of young people with higher education degrees, which is still too low in Italy.

D. Contini (✉) · R. Zotti
Department o of Economics and Statistics "Cognetti De Martiis", University of Torino, Torino, TO, Italy
e-mail: dalit.contini@unito.it; roberto.zotti@unito.it

# 1  Introduction

There is a huge literature from the economics and sociology of education analysing the role played by family background and economic resources on individuals' schooling and college choices. Overall, this body of work provides overwhelming evidence that educational choices are strongly influenced by family background. It is widely recognised that, on average, children from higher socioeconomic status backgrounds perform better at school: this pattern is attributed to the capability of more advantaged parents to purchase better quality education, offer cultural stimuli, and support their children in case of difficulties. Yet, students from advantaged backgrounds make more ambitious school choices and exhibit better outcomes net of prior scholastic results. Further differences in educational choices across family backgrounds may emerge because, acknowledging their own ability, rational individuals take decisions according to costs and expected benefits, maximising a utility function.

Breen and Golthorpe (1997) conceptualise utility in terms of expectations concerning the social class destinations of their offspring and, emphasising the role of aspirations, assume that individuals aim at minimising the risk of social demotion (i.e. ending up in a lower class than that of their parents). Parental education is also valued as a major driver of aspirations, and most empirical analyses of the effects of family background on educational outcomes either focus on the role of parental education or control for it. Other channels might exacerbate differences across family backgrounds in retention. Tinto (1975, 1993) highlights the role played by academic and social integration. Student academic performance and interaction with faculty, as well as involvement in informal peer-group interactions, may lead to either positive or negative experiences that affect feelings of inclusion. Students who feel more disconnected are more likely to withdraw: because first-generation university students often lack good knowledge of and familiarity with the higher education system, they tend to have a higher chance of experiencing poor integration and eventually drop out.

A large body of the economic literature is centred on the role played by family income, and the utility function is defined in terms of children's future earnings. As discussed in Becker (1975), low-income families may face limited borrowing opportunities. Credit constraints may discourage college attendance among youth from low-income families, even when the financial returns are high. However, Cameron and Heckman (2001) and Carneiro and Heckman (2002) find relatively small gaps by family income after controlling for children's ability. They conclude that the long-run factors associated with family income—family environment, early investments in children's education—are what play a prominent role in explaining differential college enrolment rates by family income compared to short-term

borrowing constraints. Similarly, Stinebrickner and Stinebrickner (2008) study college dropout decisions and report little evidence of credit constraints on most students. Instead, other scholars find that financial constraints are important drivers of university enrolment and completion (Ellwood & Kane, 2000; Belley & Lochner, 2007). Comparing cohorts from the mid-seventies studied in Heckman and colleagues' work with cohorts of students from the mid-nineties, Belley and Lochner (2007) find that family income has become substantially more important over time. They conclude that it is likely that borrowing constraints have become more stringent, although they acknowledge that other factors such as social networks, imperfect information and college admissions policies might have played a major role as well. Bound et al. (2012) find that growing difficulties in financing a college education, especially among students from low-income families, have contributed to increasing student employment to cover a greater share of college costs, and in turn to increasing time to degree. Examining college dropout, Stinebrickner and Stinebrickner (2012) argue that students learn about their academic ability from grade performance while in college and provide evidence that a substantial share of withdrawals can be attributed to the gained awareness of poor performance. Indeed, families invest more in their children's education the higher the expectations are of their ability (Checchi, 2000). While affluent parents might still find it worthwhile to keep financing their offspring's education even when they perform poorly, low-income ones are more likely to give up.

The issue of credit constraints is addressed mainly in research on the USA and UK, where the tertiary education system is strongly differentiated, and tuition fees are generally much higher. In European countries, where higher education institutions are mainly public and direct costs are much lower, the explanations put forward by scholars of the potential influence of family income on university attendance (conditional on prior ability and schooling careers) are more generically related to the inability to face costs, including the cost of living, and to foregone earnings (Glocker, 2011; Barone et al., 2014). Where financial difficulties and no efficient student aid system exist, disadvantaged students often need to cover their costs by working, increasing time to degree and/or leading to dropout (Glocker, 2011; Triventi, 2014). In favourable labour market conditions, pull factors may also operate, as in particular, low-income students might be induced to accept good job offers and leave university. Indirect evidence of an impact of family income on higher education attendance and completion is also provided by the numerous studies showing the beneficial effect of student aid in different countries (e.g. Dynarski, 2003; Glocker, 2011; Mealli & Rampichini, 2012; Singell, 2014; Bettinger et al., 2019; Denning et al., 2019; Modena et al., 2020).

Against this background, in this chapter, we analyse whether family economic conditions affect the probability of dropout from Italian university courses upon enrolment. Italy is an interesting study case because the education system is mainly public and university tuition fees are relatively low and income progressive. While parental education and occupation may shape aspirations—and thus the wish to undertake ambitious educational programmes—lack of income could represent a material obstacle to the continuation of study. However, because the direct costs for

disadvantaged students are low, we would expect income not to be highly relevant in this context. As we will show, this is not the case: economic conditions appear strongly associated to student dropout, even after controlling for other dimensions of socioeconomic background, prior school achievement and school type. To our knowledge, there is little existing evidence in Italy on the role played by financial conditions on student academic careers in university. One reason is the lack of appropriate data. Although administrative data provide a measure of family income, it is difficult to identify its independent effect because of the potential confounding of other family background characteristics.

Our research focuses on student educational careers upon enrolment in higher education. Exploiting a unique data set from the University of Torino that links administrative data from students' university careers, information on family income and wealth and information on mothers' and fathers' characteristics collected at matriculation, we disentangle the effects of income, parental education and parental occupation on the probability of dropping out in the first academic year. Information on the financial situation of the family is provided by the ISEE indicator (*Indicatore della Situazione Economica Equivalente*), which is an official document released by the tax authorities delivering a measure of the household economic condition, based on official records of family members' labour income, property and real estate assets, and normalised by the number of components. This document is used to determine tuition fees due for each student.

Parental education and occupation are not available in university registries. To overcome this limitation, the University of Torino has been collecting data on parental education and occupation since the 2014/2015 academic year through an online questionnaire that students fill in at matriculation. Although this section is not mandatory, the large majority (approx. 90%, evenly distributed across subgroups) provide this information. However, nearly 30% of the students do not disclose the ISEE documentation. We show that these data are not randomly missing and that a non-negligible share can be attributed to early dropout decisions. Because in this case complete case analyses or naïve solutions will deliver biased estimates of income effects, we tackle this problem by implementing an appropriate ad hoc imputation strategy.

The rest of the chapter is organised as follows. In Sect. 2, we summarise the existing evidence for Italy. In Sect. 3, we describe the data, and in Sect. 4, we illustrate the problem of missing information investing income data and how we tackle it. In Sect. 5, we describe the empirical strategy, and in Sect. 6, we present our findings. Conclusions follow in Sect. 7.

## 2 The Italian Context

Despite the absence of formal barriers to track choice and access to university, the Italian educational system is flawed by strong socioeconomic inequalities (Cobalti & Schizzerotto, 1993; Checchi & Flabbi, 2007). In comparative research,

Italy stands as a country with particularly large inequalities across parental class and education in upper secondary school choice and access to tertiary education (Jackson, 2013). Family background critically influences students' high school choices (Gambetta, 1987; Schizzerotto & Barone, 2006). Even if inequalities in access to upper secondary education have consistently declined and the share of students enrolling to the academic track has increased over time, class inequalities in track choices have not changed much (Panichella & Triventi, 2014). Horizontal segregation in high school has strong consequences on inequalities in university enrolment, as the transition rate to tertiary education varies largely across tracks (around 80% for students with a lyceum diploma, and below 30% for students with a vocational/technical diploma). Overall, there is evidence of increasing participation in higher education and slightly decreasing inequalities up to the 2000s (Argentin & Triventi, 2011; Guetto & Vergolini, 2016), but in the most recent decade, probably due to the economic crisis, transition rates have been declining and differences across high school tracks have increased, which has determined a change in the composition of the enrolled population (ANVUR, 2016).

Research on student academic careers has been limited by the lack of appropriate longitudinal data at the national level. For this reason, the existing literature on university dropout is largely based on retrospective survey data on high school graduates, periodically run by the National Statistical Institute (Cingano & Cipollone, 2007; Di Pietro & Cutillo, 2008; Cappellari & Lucifora, 2009; Ghignoni, 2017; Contini et al., 2018). This literature reports substantial differentials related to family background and shows that disadvantaged groups in terms of enrolment are also disadvantaged in terms of persistence. These groups include students who attended technical institutes and vocational schools (largely composed of students of lower socioeconomic background), although parental education and social class also influence university attendance and retention, conditional on prior schooling experience. Disadvantaged students are also less likely to enrol in a second tier, once they have obtained a bachelor's degree (Bratti & Cappellari, 2012).

Only a few studies have been based on micro-level administrative data (Belloc et al., 2010, Clerici et al., 2014, Carrieri et al., 2015, Zotti, 2016; Contini & Salza, 2020, Scagni, 2021). Because the archives on schooling and university careers are not linked together, it is not possible to study enrolment choice and consider selection effects. Moreover, a major limitation is that, while it is possible to obtain data on family income, there is no information on parental characteristics. Parental education and occupation influence individuals' aspirations and shape their expectations about future life chances. Economic conditions influence the possibility of bearing the direct and indirect costs of schooling. To disentangle these effects, data on all of these dimensions are needed.

While parental education and class strongly influence high school choices, in Italy there is no evidence of income effects at this stage (Checchi, 2000). This is hardly surprising, because schooling is free up to high school completion, and the expansion of the educational system has now made high school attendance almost universal, as nearly 85% of the young attain a high school qualification. The evidence on the role of economic resources in higher education is mixed. Analysing

a national sample in the survey on Household Incomes and Wealth, Checchi (2000) reports that family income does not seem to play a significant role in preventing the enrolment of cohabiting children in Italian public universities. Instead, Aina (2013) finds sizable effects on enrolment probability but small effects on dropout. Using administrative data from single institutions, Zotti (2016) and Scagni (2021) report income effects on dropout probability. Although analysing the data of single institutions has limited external validity, focusing on more homogeneous environments has the advantage of better controlling for contextual confounding effects. Analysing the University of Salerno, Zotti (2016) reports significant differences between low- and medium-income families in dropout probability. Scagni (2021) analyses data from the University of Torino and finds a sizable effect of income on dropout choices. Belloc et al. (2010), however, report the opposite finding—that low-income students drop out less—for the University Roma La Sapienza. Yet, this result is derived from including university performance (a mediator of dropout) as a control, and thus it is not comparable with the other studies. From a different perspective, Barone et al. (2018) use measures of material deprivation to study university enrolment and find that economic deprivation, as such, matters, even controlling for other variables meant to capture the rational choice mechanisms, in line with the Breen and Goldthorpe's theoretical model, although it does not play a major role.

Indirect evidence of the role of financial conditions on student academic careers is provided by the compelling evidence that income support provided to low-income students is effective in preventing dropout and fostering in-time graduation (Mealli & Rampichini, 2012; Vergolini & Zanini, 2015; Martini et al., 2021; Modena et al., 2020). Scholarships may favour college enrolment and persistence by providing income that allows students to allocate more time to school activities instead of work.[1]

## 3   Data

We exploit administrative data provided by the Ministry of Education on the entire career of the cohorts of students first enrolled at the University of Torino in a bachelor's programme in the three academic years from 2015/2016 to 2017/2018. The archive contains full information on the students' progression (including exam transcripts and credits earned, degree changes, timing of degree attainment or withdrawal); demographic characteristics (gender, age, place of birth and place of residence); and information on previous schooling (type of high school and final examination marks). These data have been integrated with information on

---

[1] As shown by Triventi (2014), students from upper-middle classes have a lower probability of working while studying, and working students have much poorer performance outcomes than full-time students.

family income and tuition payments, with information on scholarship recipiency[2] and with a unique piece of information on parental education and occupation collected independently by the University of Torino at matriculation since 2014.[3] This makes it possible to improve our understanding of socioeconomic inequalities in higher education, assess the independent contribution of each of these family characteristics and disentangle the effect of economic conditions.

We analyse the determinant of first-year dropout, with a particular focus on the role played by family income. Withdrawal is defined implicitly, based on whether we observe re-enrolment in year 2. Because we have access only to microdata from the University of Torino, we cannot distinguish between changes of institution and withdrawal from higher education altogether.[4] Previous analyses based on more comprehensive data have, however, shown that, among bachelor students, only a small share of the observed dropouts belong to the former group, so we believe we can safely interpret the results in terms of system-level dropout.

In Italian public universities, tuition fees are progressive, depending on household economic conditions. Students make a first payment of a fixed amount at the beginning of each academic year. In late fall, they are asked to provide the ISEE document reporting the family equivalized indicator, based on family members' labour income, properties and real estate assets.[5] Students whose ISEE exceeds a given threshold (currently set around 85,000 euros) or not providing the document are requested to pay the maximum fee (approximately 2500 euros per year). Nearly 30% of students do not provide the ISEE declaration. In the next section, we deal with this issue: as we will show, this piece of information is clearly not missing randomly. This implies that we cannot ignore the issue and conduct a complete case analysis: instead, missing data will be imputed, based on the available information on the following academic years, on parental education and occupation and tuition payments.

## 4 Missing Data on Family Income

If we could assume that, conditional on observed variables, data on income were "missing at random" (MAR), we could conduct a complete case analysis including all of the relevant explanatory variables in the models. There are, however, good reasons to believe this is not the case. First, because high-income students have

---

[2] Data on scholarships was made available by EDISU-Piemonte (Ente Regionale per il Diritto allo Studio Universitario).

[3] This data collection was spurred by the project EqualEducToEmploy, financed by the Compagnia di San Paolo in 2012–2016.

[4] Students changing their degree programme are not considered dropouts.

[5] Students may figure as an independent household only if they have lived on their own for at least 2 years and if they have earned at least 7000 euros/year. This rule was introduced in the early 2000s to discourage the previous common practice of changing residence to figure as a separate, low-income household and pay low tuition fees.
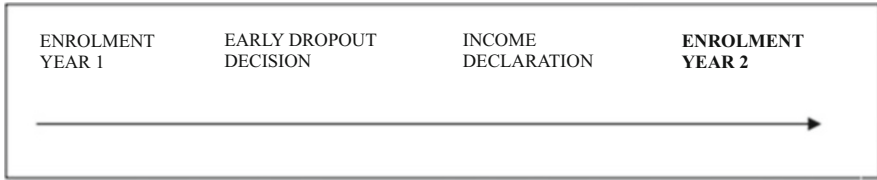
| ENROLMENT YEAR 1 | EARLY DROPOUT DECISION | INCOME DECLARATION | **ENROLMENT YEAR 2** |
|---|---|---|---|

**Fig. 1** Decision-making timeline

no tuition reductions, they have no incentive to provide an income declaration. Let us label these students *rich*. Indeed, if we could assume that all individuals with missing ISEE exceed the highest threshold, it would not be a big problem, because we would have relevant information on income that we could exploit. Unfortunately, there is evidence against this assumption. When we analyse the characteristics of the students with missing ISEE we find that: (a) many of the students with missing ISEE come from disadvantaged family backgrounds in terms of parental education and occupation (see Table 9 in Appendix A); and (b) many students not disclosing income in year 1 do so in subsequent years, often reporting a low ISEE value (see Table 10 in Appendix A). If economic conditions are fairly stable over a short time span, we may assume that in year 1 they had missed the deadlines, so we call these students *sloppy* and exploit the information provided in later years.

Second, students who decide to leave their studies within the first couple of months of the academic year also have no incentives to declare ISEE, because ISEE determines the second tuition payment, due in late fall. We call these students *early dropouts*. The choice timeline is depicted in Fig. 1.

While the *rich* and *sloppy* can be easily handled by imputing high income or subsequent ISEE values, *early dropouts* involve an endogeneity issue that must be considered. Endogeneity results from the fact that, although we are dealing with missing values for an independent variable, whether this variable is observed or not may depend on the dependent variable itself.[6] Hence, we cannot simply ignore the issue and exclude these cases from the analysis, because we would end up with potentially highly biased estimates of the effect of income on the dropout probability. As we will see later, this practice would lead to substantial underestimation of the effect of interest.

We now describe how to identify the students in these subgroups and our imputation strategy. We classify the students in the cohorts of interest in terms of whether they have or have not provided the income declaration in academic years 1 and 2, whether they have or have not enrolled in year 2 and, when relevant, whether they have paid the second tuition instalment: this piece of information is useful to identify early dropout students. Details are provided in Table 1.

---

[6] Endogeneity refers to situations in which an explanatory variable is correlated with the error term. In other words, an endogenous variable is a variable whose value is determined by the model, while an exogenous variable is one whose value is determined outside the model.

**Table 1** Classification of students (matriculated population in BA degrees, 2015–2017)

| YEAR 2<br><br>YEAR 1 | Not enrolled in year 2<br>(dropout) | | Enrolled in year 2<br>(not dropout) | | TOTAL |
| --- | --- | --- | --- | --- | --- |
| | | | ISEE<br>provided | ISEE<br>not provided | |
| ISEE<br>provided | N=3120<br>9.32% | | N=19424<br>58.01% | N=1105<br>3.30% | N=23649<br>70.63% |
| ISEE<br>not provided | **Second tuition payment** | | N=1698<br>5.07%<br>*(SLOPPY)* | N=5873<br>17.54%<br>*(RICH)* | N=9836<br>29.37% |
| | Yes | N=770<br>2.30%<br>*(OTHER DROPOUTS)* | | | |
| | No | N=1495<br>4.46%<br>*(EARLY DROPOUTS)* | | | |
| TOTAL | N=5385<br>16.08% | | N=21122<br>63.08% | N=6978<br>20.84% | N=33485<br>100% |

Note: Authors' elaboration.

Most of the students (more than 70% of the entire student population matriculated in bachelor's degree courses) provide ISEE in year 1. Consider the students not declaring ISEE in year 1 (29.37% of the total population); as discussed above, we may identify three relevant clusters: the *rich*, the *sloppy* and the *early dropouts*, as well as an additional residual group. In the following lines, we describe how we identify them and the imputation strategy. Let us start with those who do not drop out by year 2.

1. *SLOPPY*. As argued above, we assume that those who did not declare income in year 1 but declare income in year 2 had previously missed the deadlines: the *sloppy* represent 5.07% of the total population. Assuming short-term stability of economic conditions, we impute ISEE in year 1 using the value reported in year 2.

2. *RICH*. Some students fail to provide the information even in year 2 (and in subsequent years). These students (17.5% of the total population) are labelled *rich*, under the assumption that if a student does not disclose ISEE more than once, it is because there would be no substantial tuition reduction justifying the burden required to produce the documentation. For these individuals, we impute ISEE with a conventional value exceeding the maximum threshold. To keep it simple, we impute the value 100,000 and run robustness checks with alternative values (see Sect. 6).

After these imputations, the share of students with no information on economic condition drops from 29.37% to 6.76%. Even if the size of the missing ISEE population is small at this point, we must still account for the most problematic subgroup of students: those who do not enrol in year 2.

3. *EARLY DROPOUTS*. To identify this group, we exploit an additional piece of
   information: whether students have made the second tuition payment, due in late
   fall. We assume that those who did not (4.46% of the total population) have taken
   the dropout decision before the ISEE deadline. Our imputation strategy for the
   early dropout students relies on the available information on parental education
   and occupation and on the observed relation between these family background
   characteristics and ISEE. Let us define $I$ as the household economic condition
   indicator and $z$ as the vector of dummy variables describing mother's and father's
   education and occupation. Assuming a linear relation $I_i = a + bz_i + u_i$, we
   estimate model parameters, predict ISEE for given combinations of parental
   background characteristics and use the estimated $E(I|z)$ to impute missing ISEE.
   Yet, to address the endogeneity issue, we must acknowledge that the relation
   between $I$ and $z$ is generally different in the dropout population from that in the
   student population at large, because economic conditions and other dimensions
   of family background may themselves affect dropout (see proof in Appendix
   B). Against this background, we estimate the relation between $I$ and $z$ among
   those dropouts disclosing income and impute the predicted expected value
   $\hat{E}(I|z, \text{drop out})$, under the additional assumption that the same relation holds
   for early and late dropouts.
4. *OTHER DROPOUTS*. There is an additional small residual group of dropouts
   (2.3% of the total population), who did not declare ISEE in year 1, but, having
   paid the second instalment, should not be considered as early dropouts. In
   principle, we could exploit the observed relation between parental characteristics
   and income and impute expected income as for the early dropouts; however, this
   would imply neglecting their decision not to disclose their income. Instead, we
   may acknowledge that this group is likely to be composed of *sloppy* and *rich*
   students. However, because they drop out, we cannot observe their behaviour
   in year 2, so we have no means of identifying them. Hence, we will assume
   they are all rich. Although this is unlikely to be true for all of the students
   in this group, by imputing a high value of income to all of them, we tend to
   narrow the economic differences between dropouts and non-dropouts, delivering
   a conservative estimate of the true income effect.

## 5   Empirical Strategy and Variables Description

The original sample included 33,485 individuals who first matriculated in bachelor's
degree programmes between 2015 and 2017. We excluded from the analyses the
students not reporting parental occupation or parental education for both parents
(approximately 10% of the original sample, apparently randomly selected) and those
who attained a high school degree abroad, because most of them did not report
family background information (final sample size $N = 29{,}719$).

In Table 2, we show descriptive evidence on the ISEE and the parental edu-
cation distributions of dropouts and non-dropouts. On average, the former display

substantially less favourable economic conditions and a smaller share have parents with higher education degrees. In the last columns, we report the share of dropouts within the population at large and among those providing and not providing the income declaration. As we can see, dropouts are overrepresented among those not disclosing income, confirming the suspicion that provision of the income declaration may be endogenous to the early dropout decision.

To analyse the role of family economic conditions on dropout probability, we estimate logit models where the dependent variable is a binary indicator taking the value 0 if the students enrolled in year 1 re-enrol in year 2 and the value 1 if they do not re-enrol, focusing on students who first matriculated between 2015 and 2017 in 3-year degree programmes. We consider the following baseline specification:

$$D_i^* = \beta_0 + \beta_1 I_i + \beta_2 x_i + \beta_3 z_i + \beta_4 f_i + \beta_5 c_i + u_i \tag{1}$$

$$D_i = \begin{cases} 1 & if \ D_i^* > 0 \\ 0 & if \ D_i^* < 0 \end{cases} \tag{2}$$

where $D^*$ is the latent utility of dropout, $D$ is the observed binary counterpart and the error term $u$ is distributed as a logistic random variable. The explanatory variable of main interest is $I = ln(\text{income})$, while the control variables are $x = $ parental education and occupation, $z = $ socio-demographic characteristics and prior schooling, $f = $ field of study and $c = $ matriculation cohort.

Given that we can control for a large array of explanatory variables capturing all of the main determinants described in the existing literature (including other dimensions of socioeconomic background), we are able to estimate the independent effect of family economic conditions on the probability of withdrawal. What often prevents researchers from being able to interpret the income effect as causal is the unavailability of information on parental education and occupation. In the absence of such controls, due to the association between these variables and family income, we would not be able to disentangle income effects from other effects related to family background. Moreover, there are possible selection effects that might affect our results, because by observing only university students we cannot model the enrolment decision. We address these limitations in Sect. 6.3. The explanatory variables are defined as follows:

- *Income* is defined as the natural logarithm of the ISEE indicator, determining family economic conditions from household income, parental wealth and family size. When missing, we use the imputation strategy described in Sect. 4.
- *Parental education* is recorded separately for mothers and fathers, according to the following classification: up to lower secondary school, upper secondary school and higher education. However, in the estimation, we include the highest level between mother and father, further distinguishing between households where one parent or both parents have a university degree.

**Table 2** Family economic condition (ISEE) and parental education by dropout status and share of dropouts among students providing and not providing the income declaration

|  | ISEE | | | Parental education | | | | | |
|---|---|---|---|---|---|---|---|---|---|
|  | 25th percentile | 50th percentile | 75th percentile | Both ≥ high school | At least 1 higher education | Both higher education | All % | Decl. ISEE % | Not decl. ISEE % |
| Dropouts | 9886 | 18,325 | 29,025 | 75.8% | 22.2% | 6.8% | 15.5 | 12.7 | 16.7 |
| Non-dropouts | 15, 819 | 28,990 | 65,116 | 80.7% | 29.4% | 10.2% | 84.5 | 87.3 | 83.3 |

Note: Authors' elaboration

- *Mother and father occupation* is categorised as: blue collar, low-skilled white collar, high-skilled white collar and self-employed.[7] For the mother, we also add the category housework.
- *Female* is dummy variable identifying female students to account for the widespread evidence of gender differences in educational outcomes.
- *Age at matriculation* is included because there is extensive evidence that individuals not enrolling right after the end of high school (possibly after a period of occupation or while working) or, more generally, at an older age (perhaps because they previously experienced grade repetition) are more likely to leave university before degree completion. The variable is included in a categorical version (<=19 years old, 20 years old, 21–25 years old and more than 25 years old) to capture possible non-linear effects.
- *High school track* is included because prior schooling has been shown to strongly affect higher educational choice and outcomes. It is classified into traditional lyceums (classic and scientific), other lyceums (linguistic, human science, artistic), technical schools and vocational schools. Students who attended high school abroad ($n = 881$) were excluded from the analyses.
- *High school final grade*, ranging between 60 (pass) to 100 (excellent) is a proxy of academic preparedness and has been shown to be an important predictor of students' outcomes.
- *Area of origin* may influence the dropout probability for several reasons. First, because there is evidence from national and international standardised assessments that the level of competencies reached in school widely differs across the country (highest in the North and lowest in the South, see Bratti et al., 2007). Second, because students leaving their family of origin and bearing higher costs of living, on the one side are more exposed to changes in family economic conditions, but on the other side, they might be more motivated than stayers. The area of origin has been based on information related to high school location. We adopt the classification: Turin, Piedmont, North-West, North-East, Centre and South.
- *Field of study*. University careers—withdrawal/completion, credit attainment speed, grades—vary across majors and disciplines. We classify the field of study into broad categories: Scientific, Political and Social Sciences, Economics, Humanities, Health and Psychology.[8]
- *Scholarship* is a binary variable taking value 1 if the student receives financial aid in the form of a (small) scholarship and 0 otherwise. In some specifications,

---

[7] 'Blue collar' includes workers; 'Low-skilled white collar' includes clerks and service workers; 'High-skilled white collar' includes senior officials, professionals, teachers and managers; and 'Self-employed' includes business owners, self-employed and freelance.

[8] Scientific includes Mathematics, Physics and Natural sciences; Political and Social Sciences includes Law and Political Sciences; Economics includes Business, Management and Economics & Statistics; Humanities includes Philosophy, History, Languages, Arts and Educational sciences; Health includes all healthcare professions (Nursing, Speech Therapy, Physiotherapy, Dental Hygiene, etc).

we include the variable in the model to account for the evidence that financial aid
has a beneficial effect on student progression.

– *Working student* is a binary variable taking the value 1 if the student declares
  being a working student and 0 otherwise.[9] In some specifications, we include
  this variable because this condition often entails worse academic outcomes and
  higher chances of withdrawal.

Descriptive statistics on the full set of variables are presented in Table 3.

## 6  Results

In Table 4, we summarise the results of logit model estimation relative to the effect
of income on the dropout probability. All models control for parental education and
occupation, gender, age at enrolment, high-school type and final grade and area of
origin, as well as including field of study and cohort fixed effects. For comparative
purposes, we start with two naïve strategies: a complete case analysis (column 1)
and a model including all observations, with a variable taking the observed *ISEE*
value if available and 0 if missing and a dummy indicator for missing *ISEE* (column
2).[10] We then move to models using the imputed ISEE, according to the procedure
described in the previous section: a model with the baseline explanatory variables
(column 3) and models adding as control variables an indicator of the student being
a scholarship recipient and whether the individual is a working student (columns
4–6).[11]

The effect of income is negative and highly significant in all models, implying
that students from more affluent families experience lower chances of withdrawal.[12]
The effect appears weaker in the complete case model than in the models where
we address the missing data issue with appropriate imputation. The effect is
even weaker when we estimate the naïve model in column 2: interestingly, the
estimates reveal that the dropout probability for individuals not disclosing ISEE
is substantially larger even than the probability experienced by those reporting

---

[9] Working students may be eligible for part-time status, which means that they are given twice
the time to complete their degree programmes and are entitled to pay reduced tuition fees.
Unfortunately, although the administrative data report whether a student declares being a working
student, we do not know whether they apply for part-time status.

[10] In this way, the income coefficient describes the effect of ISEE among those who declared it,
and the missing ISEE dummy coefficient captures the difference between those who do not provide
ISEE and individuals with ISEE = 0.

[11] In the first year, the scholarship is granted according to family income, although only
approximately half of the eligible students apply for it. From the second academic year upon
enrolment, merit restrictions also apply.

[12] By making a single imputation for each missing ISEE value, the standard error of the estimates
will be underestimated to some extent. However, due to the large sample size, we are confident that
the estimates will still be highly statistically significant.

**Table 3** Descriptive statistics

| Variable | Obs | Mean | Std. Dev. |
|---|---|---|---|
| Dropout | 29,719 | 0.155 | – |
| Age at enrolment | | | |
| <=19 years old | 29,719 | 0.693 | – |
| 20 years old | 29,719 | 0.182 | – |
| 21–25 years old | 29,719 | 0.102 | – |
| (>25 years old | 29,719 | 0.021 | – |
| Gender | | | |
| Females | 29,719 | 0.600 | – |
| Males | 29,719 | 0.399 | – |
| Secondary school | | | |
| Lyceum | 29,719 | 0.528 | – |
| Other Lyceum | 29,719 | 0.149 | – |
| Technical | 29,719 | 0.210 | – |
| Vocational | 29,719 | 0.082 | – |
| High school missing | 29,719 | 0.029 | – |
| High school grade | 29,719 | 77.51 | 11.06 |
| Education of parents | | | |
| Lower secondary | 29,719 | 0.203 | – |
| Upper secondary | 29,719 | 0.521 | – |
| One higher education | 29,719 | 0.177 | – |
| Both higher education | 29,719 | 0.097 | – |
| Occupation father | | | – |
| Blue collar | 29,719 | 0.260 | – |
| Low skilled white collar | 29,719 | 0.238 | – |
| High skilled white collar | 29,719 | 0.139 | – |
| Self-employed | 29,719 | 0.326 | – |
| Occupation missing | 29,719 | 0.035 | – |
| Occupation mother | | | |
| Blue collar | 29,719 | 0.184 | – |
| Low skilled white collar | 29,719 | 0.350 | – |
| High skilled white collar | 29,719 | 0.154 | – |
| Self-employed | 29,719 | 0.171 | – |
| Housework | 29,719 | 0.117 | – |
| Occupation missing | 29,719 | 0.020 | – |
| Income (log) | 29,719 | 10.21 | 0.98 |
| Scholarship | 29,719 | 0.056 | – |
| Student worker | 29,719 | 0.161 | – |
| Field of study | | | – |
| Scientific | 29,719 | 0.249 | – |
| Political science | 29,719 | 0.211 | – |
| Economics | 29,719 | 0.195 | – |
| Humanities | 29,719 | 0.209 | – |

(continued)

**Table 3** (continued)

| Variable | Obs | Mean | Std. Dev. |
|----------|-----|------|-----------|
| Health | 29,719 | 0.105 | – |
| Psychology | 29,719 | 0.029 | – |
| Residence | | | – |
| Torino | 29,719 | 0.606 | – |
| Piedmont | 29,719 | 0.240 | – |
| North-West | 29,719 | 0.058 | – |
| North-East | 29,719 | 0.007 | – |
| Centre | 29,719 | 0.008 | – |
| South | 29,719 | 0.078 | – |
| Cohorts | | | |
| Cohort 2015 | 29,719 | 0.303 | – |
| Cohort 2016 | 29,719 | 0.340 | – |
| Cohort 2017 | 29,719 | 0.355 | – |

Note: Authors' elaboration

very poor economic conditions, confirming the suspicion that missing income is at least partially endogenous. In column (3), we find our preferred estimates, which we explain in further detail below. The average marginal effect (AME) is $-0.234$; thus, between the 5th and the 95th income percentile (8.45 and 11.51), the dropout probability of two otherwise identical individuals in terms of demographic characteristics, prior schooling, field of study and parental background, differs by 7.16 percentage points.[13] The effect size is large, if we consider that the overall dropout share in the first academic year is 15–16%. In columns (4)–(6) we include the additional controls: the income effect increases when we include the scholarship variable and decreases slightly when we include the variable student worker. Interestingly, the effects of both controls are large and highly significant. Ceteris paribus, scholarship recipients have a dropout probability which is approximately 8 percentage points lower than that of non-recipients: this result confirms the findings of rigorous impact evaluation studies reporting a positive impact of scholarships on student academic careers. Student workers also have a much higher dropout probability (13 percentage points) than non-workers.

We believe the overall effect of income is best captured by the model that does *not* include being a scholarship recipient and being a working student as explanatory variables (Table 4, Column 3), because these variables are endogenous to income and play the role of mediators. Both receiving the scholarship and being a working student are influenced by income: by including them in the model as controls, we would capture the direct effect of income on dropout probability, while failing to acknowledge the—positive or negative—indirect effects. Let us be more specific.

---

[13] Robustness checks with alternative values of imputed income (80,000, 100,000 and 120,000) are shown in Table 11 in Appendix A. Only marginal changes are observed.

**Table 4** The effect of economic conditions on first year dropout probability (AME)

|  | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| (log) Income | −0.0153*** | −0.0124** | −0.0234*** | −0.0302*** | −0.0215*** | −0.0276*** |
|  | (0.0041) | (0.0042) | (0.0031) | (0.0028) | (0.0026) | (0.0022) |
| Income missing |  | 0.0783*** |  |  |  |  |
|  |  | (0.0095) |  |  |  |  |
| Scholarship |  |  |  | −0.0845*** |  | −0.0779*** |
|  |  |  |  | (0.0070) |  | (0.0075) |
| Working students |  |  |  |  | 0.1384*** | 0.1343*** |
|  |  |  |  |  | (0.0073) | (0.0071) |
| Cohort FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Field of study | Yes | Yes | Yes | Yes | Yes | Yes |
| Parental occupation | Yes | Yes | Yes | Yes | Yes | Yes |
| Parental education | Yes | Yes | Yes | Yes | Yes | Yes |
| Individual characteristics | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 21,114 | 29,714 | 29,714 | 29,714 | 29,714 | 29,714 |

Note: Robust standard errors in parentheses clustered at field of study level. *** $p$-value<0.001, ** $p$-value<0.01, * $p$-value<0.05. Cohort fixed effects include 2015, 2016 and 2017. Field of study includes Scientific, Political Science, Economics, Humanities, Health and Psychology. Parental occupation includes Blue collar, Low-skilled white collar, High-skilled white collar, Self-employed for the father, and Blue collar, Low-skilled white collar, High-skilled white collar, Self-employed and Housework for the mother. Parental education includes Lower secondary, Upper secondary, Higher education or both Higher education. Individual characteristics include age (<=19, 20, 21–25 and > 25 years old), Female, High school type (Lyceum, Other lyceum, Technical and Vocational), High-school grade, Residence (Turin, North-west, North-East, Centre and South)

(1) Scholarships are typically granted to less affluent students, with the explicit aim of supporting their studies. Including the variable in the model would result in inflating the estimate of the income effect, because in this way the income effect would capture the difference in the dropout probability between more affluent and less affluent non-recipient students (or recipients, although this comparison seems less salient). In other words, in doing so we would end up interpreting the income effect as if income support policies did not exist. (2) Working students are generally less affluent than non-workers (Triventi, 2014); moreover, as we have seen, they have a much higher likelihood of leaving university before completion. By interpreting the income effect when controlling for this variable (and thus comparing students with different incomes, but either both working or both non-working), we would then end up underestimating the income effect by ascribing part of the negative effect of the lack of income to the condition of being a student worker, although being a student worker is itself influenced by the lack of income.

## 6.1 Heterogeneity of the Income Effect

Does income influence dropout probability for all students, or is the observed average effect driven by the behaviour of specific subgroups? To answer this question, we conduct separate analyses by gender, high school type, parental education, area of origin and field of study. Overall, income seems to exert a sizable influence on all subgroups, with only minor differences between them and only a few exceptions. We also estimate the income effect by the levels of the two mediator variables, indicating whether the student is a scholarship recipient or a working student. The results are shown in Tables 5, 6, 7 and 8.

Gender differences are small (Table 5). Income seems to have a slightly lower impact on the dropout probability of girls than boys, but the difference is not statistically significant. Income has a stronger effect on students holding technical and vocational high school degrees. Having previously self-selected into less academically oriented high school types, these students are likely to be more exposed to difficulties and may be able to count on lower family support than students from lyceums (Table 6, Columns 1a–4a). There are no sizable differences across parental education levels (Table 6, Columns 1b–4b). Income does not seem to exert an influence on students coming from central south Italy: we interpret this result in terms of self-selection as well. Although these students display a higher propensity to leave their studies compared to students from the North (results not presented here)—perhaps because, as shown by standardised assessments, they reach lower competence levels (Bratti et al., 2007)—they are likely to be especially positively selected in terms of aspirations and motivation and might thus be less exposed to the detrimental effects of low economic resources (Table 6, Columns 1c–4c).

Income plays a role in all fields of study except for health degrees (Table 7). This is not surprising, because of the selective admission to these programmes regulated by *numerus clausus*. Being strongly self-selected at entrance, these students are

**Table 5** Heterogeneous effects by gender

|  | (1) | (2) |
|---|---|---|
|  | Females | Males |
| (log) Income | −0.0215*** | −0.0261*** |
|  | (0.0024) | (0.0055) |
| Observations | 17,842 | 11,872 |
| Cohort FE | Yes | Yes |
| Field of study | Yes | Yes |
| Parental occupation | Yes | Yes |
| Parental education | Yes | Yes |
| Individual characteristics | Yes | Yes |

Note: Robust standard errors in parentheses clustered at field of study level. *** *p*-value<0.001, ** *p*-value<0.01, * *p*-value<0.05. Controls as in Table 4, Column 3

**Table 6** Heterogeneous effects by high school type, parental education and area of origin

|  | (1a) | (2a) | (3a) | (4a) |
|---|---|---|---|---|
|  | Lyceum | Other Lyceum | Technical | Vocational |
| (log) Income | −0.0213*** | −0.0142*** | −0.0296** | −0.0403*** |
|  | (0.0031) | (0.0029) | (0.0096) | (0.0061) |
| Observations | 15,712 | 4431 | 6254 | 2454 |
|  | (1b) | (2b) | (3b) | (4b) |
|  | Lower secondary | Upper secondary | One parent HE | Two parents HE |
| (log) Income | −0.0245*** | −0.0243*** | −0.0217*** | −0.0204** |
|  | (0.0049) | (0.0043) | (0.0051) | (0.0076) |
| Observations | 6044 | 15,510 | 5266 | 2886 |
|  | (1c) | (2c) | (3c) | (4c) |
|  | Turin | Piedmont | North | Centre-south |
| (log) Income | −0.0212*** | −0.0338*** | −0.0252* | −0.0006 |
|  | (0.0033) | (0.0026) | (0.0128) | (0.0086) |
| Observations | 18,016 | 7147 | 1968 | 2583 |
| Cohort FE | Yes | Yes | Yes | Yes |
| Field of study | Yes | Yes | Yes | Yes |
| Parental occupation | Yes | Yes | Yes | Yes |
| Parental education | Yes | Yes | Yes | Yes |
| Individual characteristics | Yes | Yes | Yes | Yes |

Note: Robust standard errors in parentheses clustered at field of study level. *** $p$-value<0.001, ** $p$-value<0.01, * $p$-value<0.05. Controls as in Table 4, Column 3

**Table 7** Heterogeneous effects by field of study

|  | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
|  | Scientific | Political science | Economics | Humanities | Health | Psychology |
| (log) Income | −0.0159*** | −0.0329*** | −0.0248*** | −0.0250*** | −0.0069 | −0.0267*** |
|  | (0.0048) | (0.0055) | (0.0045) | (0.0051) | (0.0064) | (0.0098) |
| Cohort FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Field of study | Yes | Yes | Yes | Yes | Yes | Yes |
| Parental occupation | Yes | Yes | Yes | Yes | Yes | Yes |
| Parental education | Yes | Yes | Yes | Yes | Yes | Yes |
| Individual characteristics | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 7408 | 6273 | 5800 | 6093 | 3004 | 815 |

Note: Standard errors in parentheses. *** $p$-value<0.001, ** $p$-value<0.01, * $p$-value<0.05. Controls as in Table 4, Column 3

**Table 8** Heterogeneous effects by scholarship and working student

|                          | (1a)            | (2a)                |
|--------------------------|-----------------|---------------------|
|                          | Scholarship     | No scholarship      |
| (log) Income             | −0.0172***      | −0.0325***          |
|                          | (0.0046)        | (0.0033)            |
| Observations             | 1538            | 28,039              |
|                          | (1b)            | (2b)                |
|                          | Working student | Not working student |
| Income                   | −0.0051         | −0.0253***          |
|                          | (0.0047)        | (0.0031)            |
| Observations             | 4785            | 24,929              |
| Cohort FE                | Yes             | Yes                 |
| Field of study           | Yes             | Yes                 |
| Parental occupation      | Yes             | Yes                 |
| Parental education       | Yes             | Yes                 |
| Individual characteristics | Yes           | Yes                 |

Note: Robust standard errors in parentheses clustered at field of study level. *** $p$-value<0.001, ** $p$-value<0.01, * $p$-value<0.05. Controls as in Table 4, Column 3

highly motivated and generally display very low dropout rates. Similarly, although still sizable and statistically significant, we observe a smaller income effect among students enrolled in the scientific fields, where in many degree programmes there are selective admission tests.

We find no income effects for working students, who are usually engaged in full-time jobs and display much higher dropout probabilities than full-time students. We interpret the absence of income effects for this subgroup as being related to the fact that, earning their own income, they are less dependent on family economic conditions. Income effects are weaker for scholarship recipients (AME = 0.017) than for non-recipients (AME = 0.032). This result provides additional evidence of the beneficial effect of student aid policies, as the scholarship contributes to making recipients less exposed to the negative impact of lack of family economic resources (Table 8).

## 6.2 The Effect of Parental Education and Occupation

Although the role of economic conditions emerges clearly, the effect of parental education and occupation is less clear. In Table 12 in Appendix A, we show the estimated effects for all family background dimensions. The effects of parental education go in the expected direction, but they are small and barely significant,

and even weaker results are observed for parental occupation.[14] Hence, we may conclude that at this point of the educational career—after a strong previous social selection that may be represented as an obstacle course for low-SES and a flat road for high-SES individuals—parental education and occupation do not seem to exert any substantial residual effect on the decision to complete the bachelor's degree.

## 6.3 Potential Limitations

### 6.3.1 Peer Effects

It might be argued that because we have not controlled for peer characteristics, we cannot rule out that our estimates of the effect of financial conditions also capture peer effects. Let us examine this point more closely. Students of higher socioeconomic background have, on average, better peers in terms of academic and soft skills, and better peers foster persistence in education. Hence, the link between socioeconomic background and persistence in education is likely to be causal, but (at least partly) indirect. Yet, if first-year university students' relevant peers are high school friends and classmates, it is reasonable to consider parental education and high school track—taken jointly—as good proxies for peer quality. If we believe this is the case (this is our standpoint), the issue no longer exists. If instead we believe that income as such may influence the capability of making friends and which friends young individuals make, the income coefficient might indeed also incorporate peer effects. What would the policy implications be in this latter case? If the relevant peers have been established during high school, providing financial aid upon university enrolment might not help reduce dropout, because the aid comes too late. Instead, income support could contribute to reducing dropout if the relevant peers are made after university enrolment, because this additional source of income could foster social integration in university and the acquisition of better peers. In this scenario, the income coefficient captures the total causal effect of income (direct + indirect). Thus, the policy implications may depend not only on whether the relation between economic conditions and retention is truly causal but also on the mechanisms underlying this causal link.

### 6.3.2 Self-Selection Issues

By exploiting administrative data on university students, we cannot account for selection effects related to previous educational decisions—the choice of the high school track, high school completion and university entrance. Hence, our estimates

---

[14] Even when parental occupation and ISEE are not controlled for. This result is not shown here and is available upon request.

of the effect of economic conditions on university dropout are not estimates of a causal effect in the usual sense: being conditional only on observed features, they do not capture the differences across the income distribution among individuals otherwise identical in terms of both observed and unobserved characteristics. The comparison is not fully 'like with like', because—due to the strong social selection operating along the entire schooling career—upon university enrolment, low-socioeconomic status individuals are likely to be positively selected and thus more endowed in terms of unobserved traits such as motivation and resilience than students from advantaged backgrounds (Cingano & Cipollone, 2007). For this reason, we expect our estimates to be conservative estimates of the total causal effect of income (by total effect we mean the effect inclusive of the potential effects of mediators). This conclusion holds under the assumption that motivation is independent of financial conditions after controlling for parental education and occupation (see Appendix B for proof).

## 7 Conclusions and Discussion

As maintained by Manski (1989) and more recently by Bertola (2021), college dropout need not be considered a social problem, because 'students contemplating college entrance do not know whether completion will be feasible or desirable. Hence, enrolment is a decision to initiate an experiment, one of whose possible outcomes is dropout' (Manski, 1989, p. 1). While we do agree with this point, we believe that dropout becomes a social problem if it is mainly experienced by students from disadvantaged backgrounds. If this is the case, we need to gain a better understanding of how to weaken barriers to higher education attainment among young individuals who have taken the decision to enrol in college and thus reduce intergenerational transmission of education and income.

Exploiting the unique administrative data from the University of Torino, which augments administrative university data with information on mothers' and fathers' educational level and occupation since academic year 2014/2015, we have been able to analyse whether and how family economic condition, parental education and occupation influence university students' dropout probability and disentangle their effects. We highlight the existence of a severe missing data problem, elicited by the lack of incentives to provide ISEE documentation if the student's income exceeds a certain threshold, and most importantly, in case of an early dropout decision. This source of missing data cannot be ignored. We deal with the endogenous missing data issue with an ad hoc imputation strategy and find that at this stage of the schooling career—after a strong previous social selection operating up to university enrolment—parental education and occupation no longer exert a sizable effect on educational choices. Instead, there is evidence that, despite the progressive character

of tuition fees and the existence of scholarships provided to low-income students, financial conditions have a substantial impact on university dropout.

Our results suggest that low tuition fees and current student aid policies, although beneficial, are not sufficient to eliminate the negative effect of a lack of economic resources on student academic careers. Further investigation is needed to gain a better understanding of why this is the case. While still preliminary, our analyses reveal that scholarship recipients are much less exposed to family income effects than non-recipients, even if a sizable effect also exists among them. Moreover, despite all eligible applicants receiving a scholarship in recent years, the take-up rate is low, as only about half of the students meeting the income requirements apply for a scholarship (Laudisa, 2017). Whether this is due to a lack of information or to other reasons remains to be determined, which is necessary if we wish to promote equity and at the same time raise the share of young people with tertiary education, which is still dramatically low in Italy.

## Appendix A: Additional Tables

**Table 9** Individuals with ISEE missing, father education and occupation (%)

| Occupation of father | Education of the father | | | |
|---|---|---|---|---|
| | Lower secondary | Upper secondary | Higher education | Total |
| Blue-collar | 9.31 | 3.47 | 0.19 | 12.98 |
| Low-skilled white-collar | 3.28 | 11.69 | 3.03 | 18.00 |
| High-skilled white-collar | 0.56 | 7.27 | 13.55 | 21.38 |
| Self-employed | 15.14 | 20.52 | 11.98 | 47.64 |
| Total | 28.29 | 42.95 | 28.75 | 100.00 |
| Occupation of mother | Education of the mother | | | |
| | Lower secondary | Upper secondary | Higher education | Total |
| Blue-collar | 7.08 | 3.65 | 0.26 | 10.99 |
| Low-skilled white-collar | 4.55 | 22.39 | 5.85 | 32.79 |
| High-skilled white-collar | 0.19 | 6.38 | 13.67 | 20.25 |
| Self-employed | 6.28 | 12.03 | 8.04 | 26.34 |
| Housework | 4.62 | 3.92 | 1.09 | 9.63 |
| Total | 22.71 | 48.37 | 28.91 | 100.00 |

Note: Authors' elaboration

**Table 10** ISEE distribution
in year 2

| Percentile | ISEE |
|---|---|
| 5 | 4776 |
| 10 | 8320 |
| 25 | 18091 |
| 50 | 32599 |
| 75 | 53427 |
| 90 | 81883 |
| 95 | 98306 |

Individuals with missing ISEE in
year 1 revealing ISEE in year 2
Note: Authors' elaboration

**Table 11** The effect of economic conditions on first-year dropout probability (AME)—using different values of imputed income

|  | (1) | (2) | (3) |
|---|---|---|---|
|  | 80,000 | 100,000 | 120,000 |
| (log) Income | −0.0246*** | −0.0234*** | −0.0224*** |
|  | (0.0031) | (0.0031) | (0.0031) |
| Cohort FE | Yes | Yes | Yes |
| Field of study | Yes | Yes | Yes |
| Parental occupation | Yes | Yes | Yes |
| Parental education | Yes | Yes | Yes |
| Individual characteristics | Yes | Yes | Yes |
| Observations | 29,714 | 29,714 | 29,714 |

Note: Robust standard errors in parentheses clustered at field of study level. *** $p$-value<0.001, ** $p$-value<0.01, * $p$-value<0.05. Controls as in Table 4, Column 3. Benchmark estimates in Column (2).

**Table 12** The effect of economic conditions on first-year dropout probability (AME) - All controls

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| (log) Income | −0.0208*** | −0.0199*** | −0.0239*** | −0.0234*** |
| | (0.0033) | (0.0031) | (0.0032) | (0.0031) |
| Education of the parents (ref: Upper secondary) | | | | |
| Lower sec | | 0.0069*** | | 0.0124*** |
| | | (0.0012) | | (0.0009) |
| 1 parent HE | | −0.0032 | | −0.0064* |
| | | (0.0034) | | (0.0026) |
| 2 parents HE | | −0.0053 | | −0.0113 |
| | | (0.0086) | | (0.0113) |
| Occupation of the father (ref: Low-skilled white-collar) | | | | |
| Blue-collar | | | −0.0041 | −0.0065 |
| | | | (0.0034) | (0.0036) |
| High-skilled white-collar | | | 0.0000 | 0.0033 |
| | | | (0.0091) | (0.0093) |
| Self-employed | | | 0.0174** | 0.0171** |
| | | | (0.0060) | (0.0058) |
| Occupation missing | | | 0.0011 | 0.0002 |
| | | | (0.0087) | (0.0086) |
| Occupation of the mother (ref: Low-skilled white-collar) | | | | |
| Blue-collar | | | −0.0044 | −0.0081 |
| | | | (0.0043) | (0.0045) |
| High-skilled white-collar | | | 0.0026 | 0.0064 |
| | | | (0.0075) | (0.0095) |
| Self-employed | | | 0.0169*** | 0.0162*** |
| | | | (0.0043) | (0.0048) |
| Housework | | | −0.0142* | −0.0171** |
| | | | (0.0056) | (0.0054) |
| Occupation missing | | | −0.0359** | −0.0367** |
| | | | (0.0122) | (0.0120) |
| Age (ref. <=19 years old) | | | | |
| 20 years old | 0.0278*** | 0.0278*** | 0.0279*** | 0.0281*** |
| | (0.0034) | (0.0034) | (0.0035) | (0.0035) |
| 21-25 years old | 0.0602*** | 0.0598*** | 0.0614*** | 0.0610*** |
| | (0.0117) | (0.0116) | (0.0117) | (0.0116) |

(continued)

**Table 12** (continued)

|                               | (1)         | (2)         | (3)         | (4)         |
|-------------------------------|-------------|-------------|-------------|-------------|
| >25 years old                 | 0.1781***   | 0.1747***   | 0.1834***   | 0.1784***   |
|                               | (0.0389)    | (0.0384)    | (0.0393)    | (0.0394)    |
| Female                        | −0.0240***  | −0.0246***  | −0.0234***  | −0.0240***  |
|                               | (0.0042)    | (0.0043)    | (0.0041)    | (0.0042)    |
| High school type (ref: lyceum) |            |             |             |             |
| Other-lyceum                  | 0.0677***   | 0.0670***   | 0.0683***   | 0.0673***   |
|                               | (0.0080)    | (0.0078)    | (0.0082)    | (0.0079)    |
| Technical                     | 0.0781***   | 0.0765***   | 0.0804***   | 0.0785***   |
|                               | (0.0096)    | (0.0095)    | (0.0099)    | (0.0096)    |
| Vocational                    | 0.1306***   | 0.1288***   | 0.1346***   | 0.1325***   |
|                               | (0.0062)    | (0.0058)    | (0.0071)    | (0.0065)    |
| High school missing           | −0.0363     | −0.0370     | −0.0348     | −0.0357     |
|                               | (0.0329)    | (0.0327)    | (0.0331)    | (0.0329)    |
| High-school grade             | −0.0036***  | −0.0036***  | −0.0036***  | −0.0036***  |
|                               | (0.0003)    | (0.0003)    | (0.0003)    | (0.0003)    |
| Area of origin (ref: Turin)   |             |             |             |             |
| Piedmont                      | 0.0015      | 0.0012      | 0.0005      | 0.0003      |
|                               | (0.0090)    | (0.0089)    | (0.0092)    | (0.0091)    |
| North-West                    | 0.0310***   | 0.0311***   | 0.0285***   | 0.0287***   |
|                               | (0.0062)    | (0.0063)    | (0.0061)    | (0.0062)    |
| North-East                    | 0.0554***   | 0.0577***   | 0.0520***   | 0.0558***   |
|                               | (0.0157)    | (0.0158)    | (0.0150)    | (0.0156)    |
| Centre                        | 0.1011***   | 0.1030***   | 0.0981***   | 0.1006***   |
|                               | (0.0089)    | (0.0090)    | (0.0074)    | (0.0080)    |
| South                         | 0.0651***   | 0.0652***   | 0.0653***   | 0.0653***   |
|                               | (0.0116)    | (0.0116)    | (0.0126)    | (0.0126)    |
| Field of study (ref: Scientific) |          |             |             |             |
| Political Science             | −0.0012     | −0.0013     | −0.0012     | −0.0016     |
|                               | (0.0030)    | (0.0031)    | (0.0030)    | (0.0031)    |
| Economics                     | −0.0663***  | −0.0664***  | −0.0671***  | −0.0672***  |
|                               | (0.0014)    | (0.0013)    | (0.0014)    | (0.0014)    |
| Humanities                    | −0.0107***  | −0.0111***  | −0.0105***  | −0.0111***  |
|                               | (0.0019)    | (0.0019)    | (0.0019)    | (0.0020)    |
| Health                        | −0.0955***  | −0.0958***  | −0.0951***  | −0.0956***  |
|                               | (0.0015)    | (0.0016)    | (0.0015)    | (0.0016)    |
| Psychology                    | −0.0858***  | −0.0859***  | −0.0872***  | −0.0875***  |
|                               | (0.0024)    | (0.0024)    | (0.0024)    | (0.0024)    |
| Cohort (ref: 2015)            |             |             |             |             |
| 2016                          | −0.0101*    | −0.0100*    | −0.0106*    | −0.0104*    |
|                               | (0.0050)    | (0.0051)    | (0.0050)    | (0.0050)    |
| 2017                          | 0.0115**    | 0.0116**    | 0.0115**    | 0.0118**    |
|                               | (0.0042)    | (0.0043)    | (0.0040)    | (0.0041)    |
| Observations                  | 29,714      | 29,714      | 29,714      | 29,714      |

Note: Robust standard errors in parentheses clustered at field of study level. *** $p$-value <0.001, ** $p$-value <0.01, * $p$-value <0.05

## Appendix B: Proofs

(a) The relation between income and parental education and occupation in the dropout group differs from the entire student population.

Calling $D$ the binary variable describing dropout after year 1, $I$ income and $x$ the vector of dummy variables capturing mother and father education and occupation, we now prove that:

$$E(I|x) = a + bx \neq E(I|x, D = 1) \tag{3}$$

Model (1–2) for the dropout decision assumes that the dropout probability depends on income, parental education and occupation, prior schooling characteristics, other individual variables like age at enrolment and area of residence, field of study and matriculation cohort. Simplifying the notation, we indicate with $C$ the vector of all explanatory variables other than parental education and occupation.

If $D^*$ is the latent propensity of dropping out after year 1, and

$$D_i^* = \beta_0 + \beta_1 I_i + \beta_2 x_i + \beta_3 C_i + u_i \tag{4}$$

$$P(D = 1|I, x, z) = P(D^* > 0|I, x, C) = P(u > -(\beta_0 + \beta_1 I + \beta_2 x + \beta_3 C)) \tag{5}$$

If $I = a + bx + v$, where $v$ is the error term following the usual assumptions:

$$E(I|x, D = 1) = a + bx + E(v|D = 1)$$

$$= a + bx + E(v|u > -(\beta_0 + \beta_1 I + \beta_2 x + \beta_3 C))$$

$$= a + bx + E(v|u > -(\beta_0 + \beta_1(a + bx + v) + \beta_2 x + \beta_3 C))$$

$$= a + bx + E(v|\beta_1 v > -(\beta_0 + \beta_1 a + (\beta_1 b + \beta_2)x + \beta_3 C + u)) \tag{6}$$

Even if $\rho(v, u) = 0$, the relation between $I$ and $x$ in the population of dropouts differs from that holding in the population of university students at large. The relation is weaker among dropouts because in this group $v$ is negatively correlated with $x$. If income negatively affects the dropout decision (i.e., $\beta_1 < 0$), other things being equal, individuals from advantaged parental education and occupation need a relatively low income to make the dropout choice (if income positively affected the dropout choice the opposite would hold; however, there are no theoretical reasons for this to occur).

(b) **The effect of sample selection on the estimation of the income coefficient.**

We consider the following specification for the university enrolment choice:

$$E_i^* = bSES_i + gC_i + \varepsilon_i \tag{7}$$

$$E_i = \begin{cases} 1 & if \ E_i^* > 0 \\ 0 & if \ E_i^* < 0 \end{cases} \tag{8}$$

where *SES* is socio-economic status, for simplicity defined as binary (high SES $=$ 1, low SES $= 0$) and $C$ the full array of control variables.

The dropout choice is modelled as:

$$D_i^* = \beta SES_i + \gamma C_i - u_{Mi} - u_{Li} \tag{9}$$

$$D_i = \begin{cases} 1 & if \ D_i^* > 0 \\ 0 & if \ D_i^* < 0 \end{cases} \tag{10}$$

where $u_M$ is an unobserved factor representing the individual motivation component that is not captured by the other controls such as the high school track and the final grade, and $u_L$ is the usual idiosyncratic unobserved component representing pure luck.

The causal *SES* effect is defined as the difference in the propensity to drop out between high and low *SES*, net of all individual observed characteristics and (unobserved) motivation:

$$E\left(D^*|SES = 1, C, u_M\right) - E\left(D^*|SES = 0, C, u_M\right) = \beta \tag{11}$$

Instead, the estimable effect is:

$$E\left(D^*|SES = 1, C, E = 1\right) - E\left(D^*|SES = 0, C, E = 1\right) = \beta^*$$

$$= \beta - [E\left(u_M|SES = 1, \varepsilon > -b - gC_i\right) - E\left(u_M|SES = 0, \varepsilon > -gC_i\right)] \tag{12}$$

Since $\rho(u_M, \varepsilon) > 0$ (because more motivated individuals are more likely to attend university) the expression in square parenthesis is negative, as it takes a smaller $\varepsilon$ for high *SES* individuals to enroll, and smaller $\varepsilon$ entails a smaller $u_M$. As $\beta^* > \beta$ and $\beta < 0$, $\beta^*$ will be closer to 0 (if negative) than the true causal effect $\beta$. In other words, without controlling for sample selection we will obtain an underestimate of the true (negative) effect of *SES* on the dropout probability. This argument has been made in a slightly simpler form by Cingano and Cipollone (2007).

Yet, we must acknowledge that these conclusions hold conditional on the additional hypothesis that $\rho(u_M, SES) \leq 0$. However, one might argue that higher *SES* individuals display higher aspirations and are more motivated to attain the university degree than lower *SES* individuals—to avoid social demotion, higher *SES* individuals are more prone to make ambitious educational plans (Breen & Goldthorpe, 1997). If this is true, $\beta^*$ need not be a conservative estimate of the *SES* effect, as $E(u_M | SES = 1) > E(u_M | SES = 0)$. Here, even if on average among the enrolled $\varepsilon$ is larger for low *SES* than for high *SES* (because the condition $\varepsilon > -b - gC_i$ is less stringent than $\varepsilon > -gC_i$), the expression in square parenthesis in (12) need not be negative.

On the other hand, what we are interested in here is the effect of economic conditions net of parental education and occupation. The caveats just made above should apply to the family background dimensions directly shaping educational aspirations, most likely related to the social position (parental education and social class, usually operationalized in terms of occupation) rather than to economic resources.

Against this background, our conclusion is that the effect of economic resources estimated on a sample of university students, controlling for parental education and occupation, but not accounting for sample selection, can be safely interpreted as a conservative estimate of the total causal effect of financial income on the dropout decision.

# References

Aina, C. (2013). Parental background and university dropout in Italy. *Higher Education, 65*(4), 437–456.

ANVUR. (2016). *Rapporto biennale sullo stato del sistema universitario e della ricerca*. Agenzia Nazionale di Valutazione del Sistema Universitario e della Ricerca.

Argentin, G., & Triventi, M. (2011). Social inequality in higher education and labour market in a period of institutional reforms: Italy, 1992–2007. *Higher Education, 61*, 309–323.

Barone, C., Abbiati, G., & Azzolini, D. (2014). Quanto conviene studiare? Le credenze degli studenti su costi, redditività economica e rischi di fallimento dell'investimento in istruzione universitaria. *Quaderni di Sociologia, 64*, 11–40.

Barone, C., Triventi, M., & Assirelli, G. (2018). Explaining social inequalities in access to university: A test of rational choice mechanisms in Italy. *European Sociological Review, 34*(5), 554–569.

Becker, G. (1975). *Human capital*. Chicago University Press.

Belley, P., & Lochner, L. (2007). The changing role of family income and ability in determining educational achievement. *Journal of Human Capital, 1*(1), 37–89.

Belloc, F., Maruotti, A., & Petrella, L. (2010). University drop-out: An Italian experience. *Higher Education, 60*(2), 127–138.

Bertola, G. (2021). *University dropout. Problems and solutions*. CESifo Working Paper No. 8880.

Bettinger, E., Gurantz, O., Kawano, L., Sacerdote, B., & Stevens, M. (2019). The long-run impacts of financial aid: Evidence from California's Cal Grant. *American Economic Journal: Economic Policy, 11*(1), 64–94.

Bound, J., Lovenheim, F., & Turner, S. (2012). Increasing time to baccalaureate degree in the United States. *Education Finance and Policy, 7*(4), 375–424.

Bratti, M., & Cappellari, L. (2012). Reform of higher education and social gradients in university graduates outcomes. In J. Ermisch, M. Jannti, & T. Smeeding (Eds.), *From parents to children: The intergenerational transmission of advantage*. Russel Sage.

Bratti, M., Checchi, D., & Filippini, A. (2007). Geographical differences in Italian students' mathematical competencies: Evidence from Pisa 2003. *Giornale degli Economisti e Annali di Economia, 66*(3), 299–333.

Breen, R., & Golthorpe, J. (1997). Explaining educational differentials: Towards a formal rational action theory. *Rationality and Society, 9*, 275–305.

Cameron, S. V., & Heckman, J. J. (2001). The dynamics of educational attainment for black, hispanic, and white males. *Journal of Political Economy, 109*(3), 455–499.

Cappellari, L., & Lucifora, C. (2009). The Bologna Process and college enrollment decisions. *Labour Economics, 16*, 638–647.

Carneiro, P., & Heckman, J. (2002). The evidence on credit constraints in postsecondary schooling. *Economic Journal, 112*, 705–734.

Carrieri, V., D'Amato, M., & Zotti, R. (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, 67*(4), 1034–1056.

Checchi, D. (2000). University education in Italy. *International Journal of Manpower, 21*(3/4), 177–205.

Checchi, D., & Flabbi, L. (2007). *Intergenerational mobility and schooling decisions in Germany and Italy: The impact of secondary school tracks*. IZA Discussion Papers 2876, Institute for the Study of Labor (IZA).

Cingano, F., & Cipollone, F. (2007). *University drop-out: The case of Italy*. Temi di discussione del Servizio Studi, Banca D'Italia – Eurosistema, 626.

Clerici, R., Giraldo, A., & Meggiolaro, S. (2014). The determinants of academic outcomes in a competing risks approach: Evidence from Italy. *Studies in Higher Education, 40*(9), 1535–1549.

Cobalti, A., & Schizzerotto, A. (1993). Inequality of educational opportunity in Italy, in Shavit Y. and H.P. Blossfeld Persistent inequality. In *Changing educational attainment in thirteen countries* (pp. 155–176). Westview Press.

Contini, D., Cugnata, F., & Scagni, A. (2018). Social selection in higher education. Enrolment, dropout and timely degree attainment in Italy. *Higher Education, 75*, 785–808.

Contini, D., & Salza, G. (2020). Too few university graduates. Inclusiveness and effectiveness of the Italian higher education system. *Socio-Economic Planning Sciences, 71*, 100803.

Denning, J. T., Marx, B. M., & Turner, L. J. (2019). ProPelled: The effects of grants on graduation, earnings, and welfare. *American Economic Journal: Applied Economics, 11*(3), 193–224.

Di Pietro, G., & Cutillo, A. (2008). Degree flexibility and university drop-out: The Italian experience. *Economics of Education Review, 27*(5), 546–555.

Dynarski, S. M. (2003). Does aid matter? measuring the effect of student aid on college attendance and completion. *The American Economic Review, 93*(1), 279–288.

Ellwood, D. T., & Kane, T. J. (2000). Who is getting a college education? Family background and the growing gaps in enrollment. In S. Danziger & J. Waldfogel (Eds.), *Securing the future: Investing in children from birth to college*. Sage Found.

Gambetta, D. (1987). *Were they pushed or did they jump? Individual decision mechanisms in education*. Cambridge University Press.

Ghignoni, E. (2017). Family background and university dropouts during the crisis: The case of Italy. *Higher Education, 73*, 127–151.

Glocker, S. (2011). The effect of student aid on the duration of study. *Economics of Education Review, 30*, 177–190.

Guetto, R., & Vergolini, L. (2016). *Educational expansion without equalization: Social origins and children's choice of the upper secondary track in Italy (1958-1989). No. 2016-01*. Research Institute for the Evaluation of Public Policies (IRVAPP), Bruno Kessler Foundation.

Jackson, M. (2013). *Determined to succeed? Performance, choice and education*. Stanford University Press.

Laudisa, F. (2017). *Gli studenti iscritti al primo anno che non presentano domanda di borsa di studio pur avendone potenzialmente diritto*. Osservatorio regionale per l'Università e per il Diritto allo Studio Universitario, IRES Piemonte.

Manski, C. F. (1989). Schooling as experimentation: A reappraisal of the postsecondary dropout phenomenon. *Economics of Education Review, 8*(4), 305–312.

Martini, A., Azzolini, D., Romano, B., & Vergolini, L. (2021). Increasing college going by incentivizing savings: Evidence from a randomized controlled trial in Italy. *Journal of Policy Analysis and Management, 40*(3), 814–840. https://doi.org/10.1002/pam.22260

Mealli, M., & Rampichini, R. (2012). Evaluating the effects of university grants by using regression discontinuity designs. *Journal of the Royal Statistical Society. Series A (Statistics in Society), 175*(3), 775–798.

Modena, F., Rettore, E., & Tanzi, G. M. (2020). The effect of grants on university dropout rates: Evidence from the Italian case. *Journal of Human Capital, 14*(3), 343–370.

Panichella, N., & Triventi, M. (2014). Social inequalities in the choice of secondary school: Long-term trends during educational expansion and reforms in Italy. *European Societies, 16*(5), 666–693.

Scagni, A. (2021). Performance universitaria, rischio di abbandono e background familiare: i nuovi microdati di UNITO. In M. Attanasio, M. Porcu, & G. Ragozini (Eds.), *Verso Nord. Le nuove e vecchie rotte delle migrazioni universitarie*. Franco Angeli.

Schizzerotto, A., & Barone, C. (2006). *Sociologia dell'istruzione*. Il Mulino.

Singell, L. D. (2014). Come and stay a while: Does financial aid effect retention conditioned on enrollment at a large public university? *Economics of Education Review, 23*, 459–471.

Stinebrickner, R., & Stinebrickner, T. (2008). The effect of credit constraints on the college dropout decision: A direct approach using a new panel study. *American Economic Review, 98*(5), 2163–2184.

Stinebrickner, T., & Stinebrickner, R. (2012). Learning about academic ability and the college dropout decision. *Journal of Labor Economics, 30*(4), 707–748.

Tinto, V. (1993). Leaving College: Rethinking the Causes and Cures of Student Attrition, 2nd ed., University of Chicago Press, Chicago, IL.

Tinto, V. (1975). Drop out from higher education: A theoretical synthesis of recent research. *Review of educational research, 45*(1), 89–125.

Triventi, M. (2014). Does working during higher education affect students' academic progression? *Economics of Education Review, 41*, 1–13.

Vergolini, L., & Zanini, N. (2015). Away, but not too far from home. The effects of financial aid on university enrolment decisions. *Economics of Education Review, 49*, 91–109.

Zotti, R. (2016). Should I stay or should I go? Dropping out from university: An empirical analysis of students' performances. In G. Coppola & N. O'Higgins (Eds.), *Youth and the crisis - unemployment, education and health in Europe*. Routledge.

**Dalit Contini** is Professor of Social Statistics at the University of Torino. Her current research interests cover the area of educational inequalities, higher education choices and academic careers, gender inequalities, school guidance, school systems, impact evaluation of social policies.

**Roberto Zotti** (Ph.D. University of Salerno) is Assistant Professor of Public Economics at the University of Torino. His research interests are in the field of economics of education with a focus on higher education and local economic development, and in the field of public finance with a focus on regional economics, elections, taxation and tax competition.