Effect of high school satisfaction on studying and working decisions

Advanced Microeconometrics - Final Project

Samuele Borsini | Pablo Suarez-Sucunza | Stella Vita Gatti

December 22, 2023

Abstract

High school is the place in which a society forges its future, by teaching to the youth skills useful in several environments. In this paper, we analyze the effect of being satisfied with high school on the choice of going to university, or starting to work. This choices are not mutually exclusive since a student could also both work and study or neither work nor study. Controlling for endogeneity, the study reveals that being satisfied with high school increases significantly the probability of enrolling in the university and lowers the probability of going to work right after high school graduation. Finally, the paper concludes looking for and assessing possible heterogeneity between different groups of students.

1 Introduction

The choices students make upon completing high school significantly impact their future paths and contributions to society. The quality of one's high school experience can significantly shape their future trajectory, affecting their academic pursuits and career choices.

This paper aims to study whether a high school student being satisfied or not with high school (in this case regarding the study plan of the high school) has an impact on his/her post-graduation decisions, most specifically on his/her decision to go to university or start working. These decisions are not mutually exclusive, since a student may either enroll in university, start to work, work and study at the same time, or neither work nor study. This research addresses a crucial aspect of educational and career development, shedding light on the factors that influence these decisions.

Other papers have looked at the effects of student satisfaction, like Suhre et al. (2007), which studies how it affects dropout rate at the college level, or Daily et al. (2020), which examines its importance for school absenteeism and academic performance and reports a positive

impact. However, many papers trying to study students' career choice (e.g., Dick and Rallis, 1991 and Borchert, 2002), do not account for high school satisfaction as potential contributor to this decision.

We model both the probabilities of enrolling into university and starting to work after high school, and in order to take into account the possibility that high school satisfaction is endogenous, we also model the probability that a student is satisfied or not. We do this with a trivariate probit model. We find that high school satisfaction does have a significant impact on post-graduation decisions, increasing the probability that a student goes to university, and lowering the probability that he or she starts working right after.

We also consider possible heterogeneities in the effect of high school satisfaction. For example, depending on the level of the education of the parents: the higher the level of education, the lower the magnitude of the effect on both probabilities.

The rest of this paper is structure as follows. In the next section we describe the data we use as well as the methodology we employ. Then, we discuss the main results and conclude by analyzing the heterogeneous effects.

2 Data and methodology

We use microdata from ISTAT about the "path of study and work of high school graduates" in Italy.[1] The data was published in 2015, and belongs to students who graduated high school in 2011. The interviews were conducted almost 4 years after the students had graduated. We have data on 21,795 students.

High school satisfaction. In the survey, high school graduates were asked to grade, on a scale of 1 to 10, the level of satisfaction with the study plan of their high school. From here, we created a dummy that determines if a student was satisfied if he/she answered 8 or higher, and not satisfied if he/she answered 7 or lower. We chose 8 as the threshold as to have the most even split between satisfied and non-satisfied (42.44% satisfied, 57.56% not, see figure 1).[2]

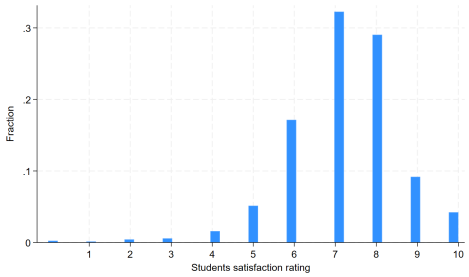

Figure 1: Distribution of level of high school satisfaction

[1] https://www.istat.it/it/archivio/96042

[2] Results, to be presented later, are robust to changing the threshold of satisfaction from 8 to 7.

Studying and working status. Both are denoted with dummies. For the studying case, the dummy equals 1 if the student enrolled in university after graduating high school, and 0 if not. For the work case, the dummy equals 1 if the individual had a job contract in 2012, that is, the year after the completion of high school. These two status are not mutually exclusive, as it could be the case that an individual was both working and studying, or also neither working nor studying. Table 1 shows the frequency of these 4 possible occurrences

Table 1: Frequencies of high school graduates working enrolled in university (total=21,795)

	Not Work	Work	Total
Not study	23.37%	21.32%	44.69%
Study	46.00%	9.31%	55.31%
Total	69.37%	30.63%	100%

Other covariates. Both mother's and father's highest level of education (elementary, middle, and high school, university degree, and post-graduate studies [3]), the type of high school that the student attended ("Liceo", "Tecnico", or "Professionale" [4]), dummies indicating whether the high school was public or private, the gender of the student, and a dummy indicating whether the student had Italian nationality or not. Finally, the final grade of the student, a dummy indicating whether the student ever failed a subject, and a dummy denoting whether the student ever changed the type of high school.

Table 2 shows the summary statistics of the variables.

2.1 Empirical strategy

We want to model the probabilities of both going to work and studying. Given that these probabilities are likely correlated (see again table 1), we propose a bivariate probit.

We assume the classical latent variable structure:

$$u_i = \begin{cases} 1 & \text{if } y^*_{ui} = s_i\alpha_u + \boldsymbol{x}_{1i}\boldsymbol{\beta}_u > \varepsilon_{ui} \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

$$w_i = \begin{cases} 1 & \text{if } y^*_{wi} = s_i\alpha_w + \boldsymbol{x}_{1i}\boldsymbol{\beta}_w > \varepsilon_{wi} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where y^*_{ui} and y^*_{wi} are the latent variables that influence the decisions of the graduated students, u_i and w_i are the Bernoulli random variables that represents whether a student chooses or not, respectively, to enrol in a university and to start working, s_i is the Bernoulli random

[3] High school is used as the base category.

[4] "Liceo" is the most university oriented type of high school, since it provides a broad and general education; "Professionale" is the most work oriented type of high school, since it provides practical skills and training for specific professions; "Tecnico" is somewhere in the middle between the other 2 types, giving both theoretical education and practical skills for specific professions.

Table 2: Summary statistics

	Mean	SD	Min	Max	N
uni_ins	.5531085	.4971829	0	1	21795
work2012	.3063088	.4609701	0	1	21795
hs_satisfied	.4283551	.4948517	0	1	21795
female	.5643496	.4958532	0	1	21795
italian	.979766	.140803	0	1	21795
public_school	.9601285	.1956618	0	1	21795
hs_professionali	.3075017	.4614696	0	1	21795
hs_tecnici	.2101858	.4074498	0	1	21795
hs_liceo	.4823125	.4996985	0	1	21795
changed_hs	.1274604	.3334957	0	1	21795
grade	76.20615	11.57859	60	101	21795
ever_failed	.19844	.3988344	0	1	21795
mother_elementary	.0684561	.2525327	0	1	21795
mother_middle	.3940353	.4886537	0	1	21795
mother_hs	.4323928	.4954195	0	1	21795
mother_uni	.023721	.1521822	0	1	21795
mother_postgrad	.0813948	.2734468	0	1	21795
father_elementary	.0819913	.2743577	0	1	21795
father_middle	.4072494	.4913333	0	1	21795
father_hs	.4028906	.4904904	0	1	21795
father_uni	.0169305	.129014	0	1	21795
father_postgrad	.0909383	.2875279	0	1	21795

variable that denotes whether the student is satisfied or not with high school, $\boldsymbol{x}_{1i}$ denotes a set of covariates that includes mother's and father's education level, the type of high school, and whether the school was public or private, $\boldsymbol{x}_i = (s_i, \boldsymbol{x}_{1i})$. We assume that ε_{ui} and ε_{wi} follow a bivariate standard normal distribution.

We also estimate the correlation across the errors

$$\rho_{uw} \equiv E[\varepsilon_{ui}\varepsilon_{wi}]$$

If the errors of the two equations were uncorrelated, we could have estimated the population parameters by running two separate probits.

Accounting for endogeneity. In the bivariate case we face the problem that high school satisfaction may be endogenous. Indeed, the satisfaction is likely to be correlated with the skill level of a student, that surely shows a relation with the choices we want to model as well, and since this measure in not included in the bivariate model, we may have endogeneity problems. To account for this, we will run a trivariate probit, modelling also the probability that an individual is satisfied with high school. Therefore, we add to the system in (2) a third equation:

$$s_i = \begin{cases} 1 & \text{if } y^*_{si} = \boldsymbol{x}_{1i}\boldsymbol{\beta}_s + \boldsymbol{m}_i\boldsymbol{\gamma}_s > \varepsilon_{si} \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

where $\boldsymbol{m}_i$ is a set of exogenous regressors used as instruments to control for the endogeneity of high school satisfaction: the final graduation grade of the student, a dummy indicating whether the student ever changed the type of high school, and another dummy denoting whether the student ever failed a subject. The data support the idea of these instruments being relevant (all of their coefficients in the third equation are significant, as shown in table 4). We assume that the instruments are also exogenous. Even if they are surely correlated with the dependent variables we want to model, we believe that this correlation runs only through the satisfaction, since a satisfied student will probably choose to enrol in a university and not to start working independently of his or her grades and high school career.

In the trivariate framework, we have also other two parameters to estimate. Let ρ_{us} and ρ_{ws} denote, respectively, the correlations of ε_u and ε_w with ε_s. These parameters signal the presence of endogeneity, therefore their significance can justify the use of a trivariate probit.

Estimation. While in estimating the bivariate probit we do not encounter any problem, we know that for the trivariate case the first order conditions for the maximization of the log-likelihood function do not have a closed form solution. Therefore, we resort to simulation based methods. To achieve consistency, the maximum simulated likelihood estimator needs a large number of draws. We use 1500 random draws in the first trivariate estimation. Then we lower this number to 1000 in the heterogeneous effects' analysis.

3 Main results

In this section we present the main results of our empirical analysis. We start by briefly commenting the bivariate model, and then go on to focus, with more detail, on the trivariate one.

3.1 Bivariate probit

Table 3 presents the estimated marginal effects of high school satisfaction on the probability of working and enrolling in university from the bivariate probit. [5]

Results indicate that the probability that a student goes to university after graduation increases by approximately 3.1 percentage points if he/she is satisfied with high school, and this increase in statistically significant. For the case of going to work, being satisfied with high school decreases the probability by 1.4 percentage points, but this effect is only significant at the 10% level.

Looking at the estimated value of ρ_{uw}, and at its t-statistic, we see that there is a significant negative correlation between the residuals of both equations, signalling that it does make sense to model these binary outcomes jointly.

Table 3: Marginal effects of high school satisfcation on the probability of studying and work

	(1) Study	(2) Work
hs_satisfied	0.0310***	-0.0146*
	(5.41)	(-2.39)
Obs	21795	
ρ	-0.399	
$\chi^2_1(\rho = 0)$	1020.9	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

3.2 Controlling for endogeneity

We now focus on the possibility that high school satisfaction is endogenous. Table 4 shows the estimates of the trivariate probit for the variable of interest.[6] We will comment the estimated values of ρ_{us} and ρ_{ws}, and then analyze the effect of high school satisfaction in this model.

As was the case in the bivariate model, the estimated ρ_{uw} is negative and significant, although now is bigger in magnitude. This signals once again that it makes sense to model both probabilities together. Both ρ_{us} and ρ_{ws} are statistically significant. They are negative and positive respectively. This confirms that high school satisfaction was endogenous when modelling both probabilities, and therefore it makes sense to model all 3 together.

[5] Marginal effects for all variables are found in the appendix.

[6] Full estimates can be found in the appendix.

Table 4: Trivariate probit coefficients for hs_satisfied, controlling for endogeneity.

	uni_ins	work2012
hs_satisfied	1.4624***	-0.7238***
	(0.0147)	(0.0503)
ρ_{uw}	-0.5027***	
	(0.0181)	
ρ_{us}	-0.926***	
	(0.0062)	
ρ_{ws}	0.4444***	
	(0.0315)	
Obs	21,795	

Standard errors in parenthesis

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5 presents the marginal effects of high school satisfaction on both the probabilities of working and going to university.[7] A student being satisfied with high school increases, on average, the probability on enrolling into university by 47.06 percentage points, and decreases the probability of starting to work by approximately 23.3 percentage points. Both these effects are significant at the 1% level.

Table 5: Marginal effects of high school satisfaction on probability of university enrollment and work

	University	Work 2012
hs_satisfaction	0.4706483***	-0.2335497***
	(0.0004018)	(0.0002988)

Bootstrap standard errors in parenthesis

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

These effects, before controlling for endogeneity were estimated to be approximately 3 and 1.4 percentage points respectively. Thud, controlling for endogeneity leads to a substantial increase in their magnitude.

4 Heterogeneous effects

In the previous section we showed that high school satisfaction does have a big effect on a student's probability to continue studying or to start working. In this one, we want to see if the magnitude and significance of these effects vary depending on the characteristics of the students, focusing on three differences: sex of the student, the type of high school the student

[7] The STATA command used to estimate the multivariate probit, *mvprobit*, does not provide a routine for obtaining the marginal effects, so they are computed "by hand", and the standard errors are obtained through boostrapping.

attended, and the level of education of the parents. Table 6 shows the sample conditional probabilities for these different groups. Such large differences across groups makes us think that there could also exist differences in the effect of high school satisfaction.

Table 6: Sample conditional probabilities of work and university enrollment for different subgroups.

		University	Work	N
Overall:		0.553	0.306	21,795
Sex of the student:				
	Male	0.472	0.361	12,300
	Female	0.615	0.264	9,495
Type of high school:				
	Liceo	0.795	0.186	10,512
	Tecnico	0.481	0.384	6,702
	Professionale	0.222	0.444	4,581
Parents education:				
	Both university or higher	0.874	0.13	1,150
	At least one lower than university	0.535	0.316	20,645

Table 7 shows the marginal effects of high school satisfaction on the probabilities of working and studying for different representative students.

Table 7: Heterogeneity analysis: marginal effects on the probability of work and university enrollment

		University	Work	N
Overall:		0.4706483***	-0.2335497***	21,795
Sex of the student:				
	Male	0.4716***	-0.2421***	12,300
		(0.00073)	(0.0005)	
	Female	0.4691***	-0.224***	9,495
		(0.00049)	(0.00035)	
Type of high school:				
	Liceo	0.4269***	-0.2325***	10,512
		(0.00065)	(0.00036)	
	Tecnico	0.5076***	-0.2261***	6,702
		(0.00038)	(0.00022)	
	Professionale	0.5253***	-0.2204***	4,581
		(0.00038)	(0.0007)	
Parents education:				
	Both university or higher	0.3272***	-0.0809***	1,150
		(0.0021)	(0.00064)	
	At least one lower than university	0.4787***	-0.2416***	20,645
		(0.00035)	(0.00027)	

Bootstrap standard errors in parenthesis

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Sex of graduate. The estimated effects on the probability of university enrollment are very similar between male and female students, the difference is of 0.25 percentage points. The difference of the effects on the probability of working is bigger: the effect on male students is about 1.8 percentage points higher in magnitude.

Parents education level. For students whose parents have a university degree or post-graduate the effects of high satisfaction on both probabilities are much smaller in magnitude, though still significant. This comes from the fact that the sample distribution of students who have both parents with university degree or higher educational level is skewed towards enrolling in university and not going into work a lot (as shown by the values reported in 1). Therefore, the high school satisfaction plays a smaller role in this case, because these students will go to university with very high probability anyway.

We can give several explanations for this behaviour: this may be due to the higher wealth of an educated family, that can give to their child the chance of enrolling easily; another possible reason is that children of higher educated people are more likely to perceive university as the natural next step in the educational path; lastly, the comparison with the parents may put pressure on the students, which then see university as necessary to not fail.

Type of high school. We notice a substantial difference between the marginal effect of high school satisfaction on the university enrollment for Liceo's students and Tecnico's one (about 8 percentage points) and between the marginal effect of satisfaction for Liceo's students and Professionale's one (about 10 percentage points). There is also a difference between Tecnico and Professionale, yet it is smaller in magnitude with respect to the others (about 2 percentage points). These differences might come from the possibility that a student who graduated in a Tecnico and a Professionale has easier access in the job market than a student from a "Liceo", therefore in this last type students are more likely to enrol in a university even if they are not satisfied with the education world, whereas in the other cases a student who is not satisfied can go to work.

We cannot say the same thing about the effect of high school satisfaction on work, since the marginal effects have about the same magnitude. Yet, there are differences, even if small, and the reasons behind them are the same seen before, yet for the job market rather than for university.

5 Conclusions

We have seen that high school satisfaction is a big factor in the decision of whether to work or to study made by high school graduates. Moreover, this effect is bigger in for individuals who by default have a lower baseline probability of going to university.

Government and policy makers, in attempting to increase education levels, should have in mind that high school satisfaction is an important determinant of the future of the students. The

policy implications of our analysis are simple: by making the high school a better place (i.e. a place that better satisfies the students), it is possible to increase the education level, and the positive effects of this phenomenon in the whole society are well known and documented.

6 Potential drawbacks of the analysis

This analysis shows significant results. Yet, there may be problems with the models and the inference.

The first is that the data include only graduated high school students, therefore the inference we made is not applicable to students who have dropped high school, because school dropout is probably correlated with high school satisfaction. Hence, we have a sample selection bias, and to overcome this problem we should use data on the students who did not graduate, yet the ISTAT dataset does not include this kind of data.

Another problem may come from the possible misspecification. Assuming linearity of the latent variables model and normality of the residuals are obviously very strong assumptions. Therefore the point estimates we did are likely to be wrong. Yet by interpreting the methods we used as approximations, rather than the exact DGP, we can interpret the magnitude and the sign of our results as similar to the true ones.

Bibliography

Suhre, C. J. M., Jasen, E. P. W. A., & Harskamp, E. G. (2007). Impact of degree program satisfaction on the persistence of college students. *Higher Education*, *54*(2), 207–226. Retrieved November 13, 2023, from http://www.jstor.org/stable/29735107

Daily, S. M., Smith, M. L., Lilly, C. L., Davidov, D. M., Mann, M. J., & Kristjansson, A. L. (2020). Using school climate to improve attendance and grades: Understanding the importance of school satisfaction among middle and high school students. *Journal of School Health*, *90*(9), 683–693.

Dick, T. P., & Rallis, S. F. (1991). Factors and influences on high school students' career choices. *Journal for research in mathematics education*, *22*(4), 281–292.

Borchert, M. (2002). Career choice factors of high school students.

Appendix A

Table 8 the coefficients of all variables forn the bivariate probit model.

Appendix B

Table 9 the coefficients of all variables forn the multivariate probit model.

Table 8: Marginal effects on the probability of studying and work

	(1) Study	(2) Work
hs_satisfied	0.0310***	-0.0146*
	(5.41)	(-2.39)
public_school	0.0316*	-0.0000325
	(2.20)	(-0.00)
hs_professionali	-0.413***	0.203***
	(-80.05)	(28.82)
hs_tecnici	-0.208***	0.153***
	(-29.73)	(19.04)
father_elementary	-0.102***	0.0510***
	(-8.87)	(4.23)
father_middle	-0.0608***	0.0391***
	(-9.41)	(5.55)
father_uni	0.0436	-0.0398
	(1.83)	(-1.52)
father_postgrad	0.115***	-0.0886***
	(9.05)	(-6.59)
mother_elementary	-0.139***	0.0169
	(-11.15)	(1.31)
mother_middle	-0.0886***	0.0338***
	(-13.87)	(4.80)
mother_uni	0.0610**	-0.0233
	(2.86)	(-1.05)
mother_postgrad	0.0690***	-0.0391**
	(5.30)	(-2.84)
female	0.0547***	-0.0471***
	(9.23)	(-7.52)
italian	-0.0333	-0.00540
	(-1.58)	(-0.26)
Obs	21795	21795
ρ	-0.399	-0.399
$\chi^2_1(\rho = 0)$	1020.9	1020.9

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 9: Trivariate probit coefficients, controlling for endogeneity.

	uni_ins	work2012	hs_satisfied
hs_satisfied	1.4624***	-0.7238***	
	(0.0147)	(0.0503)	
public_school	0.2020***	-0.06	-0.3044***
	(0.0421)	(0.0477)	(0.044)
hs_professionali	-0.9070***	0.5548***	0.0249
	(0.0221)	(0.02336)	(0.0218)
hs_tenico	-0.4412***	0.4189***	-0.0193
	(0.0227)	(0.0252)	(0.0239)
father_elementary	-0.1879***	0.1232***	-0.0452
	(0.0339)	(0.0356)	(0.0353)
father_middle	-0.1392***	0.1107***	0.0202
	(0.0191)	(0.021)	(0.0201)
father_uni	0.1173*	-0.1214	-0.026
	(0.0687)	(0.0793)	(0.071)
father_postgrad	0.2735***	-0.2578***	-0.0602*
	(0.0354)	(0.0399)	(0.0345)
mother_elementary	-0.4137***	0.093**	0.01829***
	(0.0361)	(0.0381)	(0.0377)
mother_middle	-0.2518***	0.1178***	0.1031***
	(0.0191)	(0,021)	(0.0202)
mother_uni	0.1501***	-0.0764	-0.1043*
	(0.0586)	(0.0658)	(0.0593)
mother_postgrad	0.1471***	-0.10***	-0.0211
	(0.0364)	(0.0408)	(0.0359)
female	-0.0012	-0.0743***	0.1391***
	(0.0174)	(0.0195)	(0.0182)
italian	0.4112	-0.0701	0.2359***
	(0.0623)	(0.0642)	(0.064)
ever_failed			-0.1953***
			(0.0189)
changed_hs			-0.0549**
			(0.0224)
grade			0.022***
			(0.0002)
ρ_{uw}	-0.5027***		
	(0.0181)		
ρ_{us}	-0.926***		
	(0.0062)		
ρ_{ws}	0.4444***		
	(0.0315)		
Obs	21,795		

Standard errors in parenthesis

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$