TITLE: The Effects of Private Schooling in University Acceptance in Catalonia

AUTHOR: Andrea Dana Klein Villalba

DEGREE: Business Administration and Management

ADVISOR: Katerina Chara Papioti

DATE: June 7th, 2019

Abstract

Hello, it’s me,

CONTENTS

1. INTRODUCTION

During the school-year 2016/2017, 34.5% of all secondary school students in Catalonia received a form of private schooling. The previous year, aroung 8.7% of students attended a private institution. Considering that there are over half a million secondary school students, this suggests a high demand for private schooling. University education follows a similar pattern, in which roughly 11% of students attend a private university.
Previous research [1] (Scheper, 2013) reflects on the fact that private schools in the USA are generally regarded as better than their state counterparts, and that parents may feel that they are offering a better education to their children in this way. Other researchers, look into how private schooling affects the performance of public schools [2] (Marlow, 2010), or the differences in cognitive ouctomes according to the type of school attended [3] (Coleman, Hoffer and Kilgore, 1982). Moreover, there is also reaseach on the use of SAT scores to compare schools [4] (Fetler, 1991). Further research goes into how high school drop out rates are affected by higher university enrollment rates [5](Bedard, 2001), and the demand for private schooling under different institutional arrangements [6](Stiglitz, 1974).
However, there is little research on the validity of the premise that attending a private school will increase the chances of accessing university or having better opportunities after studying. This research project seeks to determine whether or not attending a private school has any effect on university acceptance in Catalonia. Thus, for this project, we have gathered data from over 5,000 adults who studied in Catalonia between grades 1 and 12. The resulting sample, is divided between individuals born before and after 1981, when the Spanish education system was last changed as a whole. From this data set, we have done several estimations to assess the effects of different environmental factors into university acceptance.

2. PREPARING THE DATA

To be able to make estimations about the Catalan population, it was necessary to first collect the data from individuals who studied in Catalonia. Then, the data was cleaned and prepared for its use in statistical software.

2.1. Gathering the Data

To collect the data, we prepared a survey in which we asked individuals about several factors that were thought to be relevant. The survey with all the questions can be found in Appendix 1. We want precise information on the individual's education, but also general information about their family and economic situation. We do find some measurements difficult to get, since we are relying on the individuals' capacity of remembering the situation about their childhood. Thus, when it comes to family earnings, we need to take some approximate measure of the economic situation. To do so, we ask the individual about their perception of the family's situation during their childhood, and then we ask for the postal code where they lived in.

Once the survey was ready, we started spreading it to get answers. To try to ensure a correct sampling of the population, the survey was published in cities' and towns' Facebook groups [1] The criteria to choose the groups where it would be published was for the town or city to have at least 10,000 inhabitants (in some cases 9,000 if the latest data was a few years old), and then look for groups which had at least (or close to) 1,000 members. If there were several groups fulfilling these characteristics, we would try to publish in as many of them, especially for large cities. However, we tried to get the publication in at least two groups for the same city.

We requested access to 373 groups around Catalonia. Some of them required that we explain why we want to access the group, in which case we explained the purpose of the survey to the administrators. Most of the groups accepted our application. In some cases, the publication itself had to be approved by the administrator, and in most cases it was. However, a few groups eliminated the publication. Eventually, we were able to collect 6,561 answers over the course of three weeks. From these, 82.8% were female, and 83.4% had finished studying.

2.2. Cleaning the Data

When preparing the survey, we were aware that we needed a sizeable sample (300-600 answers), and we were aware of the difficulty that others have faced in collecting it. Hence, we prepared the survey expecting a sample that was a tenth of the one we obtained. This

[1] In Spain, it is popular to have and participate in Facebook groups for the town, city or neighborhood people live in.

meant that some questions that had a lot of possible answers -i.e. birth year, number of siblings, parents age, postal code, etc - were left as open questions. Thus, cleaning the data became a much larger task than initially expected.
To effectively do so, we used R, since it was the most efficient alternative. With such a large sample it may have been tempting to just eliminate those who did not answer in the expected way, but after serious consideration, it was better to try and maintain as many as possible to avoing a sample bias. The questions that required the most detail cleaning were those about birth order, parents' age when they had the first child, and postal code. Furthermore, given the amount of answers, we decided to leave untouched the questions about the parents' job and the bachelor's degree individuals accessed if they were accepted to university.
It was necessary to eliminate any answer that was not plausible -such as too many siblings, parents are too young or too old-, those that were too young -born after the year 2000- and those that did not study in Catalonia. Once we cleaned the data, we were left with 5,918 answers.

2.3. Preparing the Data

Over the cleaned data, we continued working in R to dummify the necessary variables. That is, we created dummies for the parents study level, their housing situation, and other factors necessary for our measurements. Moreover, we used the postal code to determine in which city or town the person grew up. With that, we were able to relate their city or town with the average income in 2016 for these areas. Because Barcelona is a lot larger than any other city in Catalonia, we decided to create another variable where individuals from the city were related to the average income in the neighborhood instead of the whole city.
From here, we moved on to the statistical software we would later use to make the estimations. Here we created interaction terms between several factors, and we separated the sample between individuals born before and after 1981.

3. GENERAL RESULTS

We are particularly interested in the group of individuals born after 1981 -those who studied with the current education system in Spain. However, it is interesting to see how these results change for the population as a whole and for the older individuals (born before 1982). It is important to note that, in the younger group, 48.6% of the subjects did access university, while in the older group, that was true for only 29.3% of them.

Given that we want to regress on *accessed*, we will use a probit model where the fitted value, $\widehat{accessed}$, is a continuous variable. If $\widehat{accessed}$ is above 0.5, we predict that that individual does access univerisity, and that they do not otherwise. After considering several different models, the one that fits best the objective of this research will be the one in equation 1.

$$\begin{aligned}
accessed_i = \beta_0 + \beta_1 \cdot gender_i + \beta_2 \cdot siblings_i + \beta_3 \cdot m_first_i + \beta_4 \cdot public_i \\
+\beta_5 \cdot concerted_i + \beta_6 \cdot private_i + \beta_7 \cdot extra_help_i + \beta_8 \cdot sport_i \\
+\beta_9 \cdot not_married_i + \beta_{10} \cdot m_accessed_i + \beta_{11} \cdot f_accessed_i \\
+\beta_{12} \cdot tarragona_i + \beta_{13} \cdot girona_i + \beta_{14} \cdot lleida_i + \beta_{15} \cdot early_work_i \\
+\beta_{16} \cdot eco_proxy_2_i + \beta_{17} \cdot eco_proxy_3_i + \beta_{18} \cdot eco_proxy_4_i \\
+\beta_{19} \cdot eco_proxy_5_i + \beta_{20} \cdot eco_proxy_6_i + \beta_{21} \cdot l_NB_i + u_i
\end{aligned} \tag{1}$$

Where:

- *accessed*: 1 if the individual has accessed university.
- *gender*: $1 = female$
- *siblings*: number of siblings the individual has. If the individual is an only child, this number is 0.
- *m_first*: age at which the mother had her first child.
- *public*, *concerted* and *private*: number of courses the individual did at a public, concerted or private school respectively, and $\forall i \in (0, 12)$.
- $extra_help$: whether the individual received tutoring, reinforcement, extracurricular classes or psycopedagogic assistance during at least a year.

- *sport*: whether the individual practiced sport outside school for at least a year.
- *not_married*: if the parents of the individual where not in a stable couple.
- *m_accessed* and *f_accessed*: whether the mother and the father respectively accessed university.
- *tarragona*, *girona* and *lleida*: whether the individual lived in Tarragona, Girona or Lleida respectively.
- *early_work*: 1 if the individual started working before age 18.
- *eco_proxy_j* :in reference to the perceived economic situationof the individual and $j \in (1,6)$.

  - $j = 1$: difficulty to cover basic needs.

  - $j = 2$: could just cover basic needs.

  - $j = 3$: could afford ocasional luxuries.

  - $j = 4$: could afford regular luxuries.

  - $j = 5$: could afford even more luxuries.

  - $j = 6$: money was not a limiting factor.
- *l_NB*: logarithm of the average income in 2016 in the neighborhoods in Barcelona, and towns everywhere else.

As earlier mentioned, we separated the sample in two groups. After dropping the incomplete answers, the model with the full sample uses 5817 observations. Likewise, the moder for individuals born before 1982 uses 3281 observations, and the one for individuals after 1981, uses the remaining 2536. In table 1, we can observe the estimations for the model with different samples, with their standard error and significance level. Before analyzing the effect of different variables in university acceptance, it is interesting to analyze the effectiveness of the models.
Looking into table 2, we see that the one that predicts the most observations correctly is the one for individuals born before 1982. However, looking at table 3, we see that the model mostly predicts correctly those that did not go to university. This is consistent with

the results for the full sample, even though this one does predict correctly more observations where the individual accessed university. Thus, looking at tables 2 and 3 again, we see that for individuals born after 1981 the percentage of correctly predicted observations is lower, but its distribution is a lot more uniform. While it is not precise, it does predict that roughly half of the sample would access university, which is consistent with the percentage of individuals that actually did.

An example of reference is [**?**].

4. REFERENCES

[1] Scheper, E. (2013). Comparing public and private schools.

[2] Marlow, M. L. (2010). The influence of private school enrollment on public school performance. Applied Economics, 42(1), 11-22.

[3] Coleman, J., Hoffer, T., & Kilgore, S. (1982). Cognitive outcomes in public and private schools. Sociology of education.

[4] Fetler, M. E. (1991). Pitfalls of using SAT results to compare schools. American Educational Research Journal, 28(2), 481-491.

[5] Bedard, K. (2001). Human capital versus signaling models: university access and high school dropouts. Journal of Political Economy, 109(4), 749-775.

[6] Stiglitz, J. E. (1974). The demand for education in public and private school systems.

LIST OF FIGURES

LIST OF TABLES

Appendix 1: Survey

Table 1: Model 1 with Different Sample Restrictions

Variable	**Full sample**	**Born before 1982**	**Born after 1981**
const	-1.4521 (0.1759) $***$	-1.0657 (0.254102) $***$	-1.7182 (0.2577) $***$
gender	0.0152 (0.0465)	-0.0728 (0.0635)	0.0880 (0.0700)
siblings	-0.0691 (0.0140) $***$	-0.0475 (0.0159) $***$	-0.0552 (0.0299) $*$
m_first	0.0259 (0.0042) $***$	0.0122 (0.0059) $**$	0.0417 (0.0064) $***$
public	0.0370 (0.0046) $***$	0.0232 (0.0064) $***$	0.0380 (0.0070) $***$
concerted	0.0484 (0.0044) $***$	0.0407 (0.0060) $***$	0.0539 (0.0069) $***$
private	-0.0019 (0.0060)	0.0062 (0.0075)	-0.0054 (0.0108)
extra_help	-0.0388 (0.0361)	-0.0540 (0.0492)	-0.1083 (0.0552) $*$
sport	0.1390 (0.0377) $***$	0.0121 (0.0497)	0.2242 (0.0599) $***$
not_married	-0.0806 (0.0576)	-0.1525 (0.0949)	-0.1529 (0.0750) $**$
m_accessed	0.5359 (0.0859) $***$	0.5553 (0.1469) $***$	0.4357 (0.1083) $***$
f_accessed	0.3874 (0.0715) $***$	0.4563 (0.1037) $***$	0.2826 (0.1002) $***$
tarragona	0.2341 (0.1104) $**$	0.0903 (0.1646)	0.3394 (0.1545) $**$
girona	0.0712 (0.0681)	0.0773 (0.0985)	0.0639 (0.0962)
lleida	0.1538 (0.1154)	0.0582 (0.1688)	0.2366 (0.1619)
early_work	-0.2330 (0.0357) $***$	-0.3385 (0.0501) $***$	-0.1762 (0.0527) $***$
eco_proxy_2	0.0057 (0.1047)	0.1700 (0.1579)	-0.0431 (0.148591)
eco_proxy_3	0.1969 (0.1031) $*$	0.3377 (0.1575) $**$	0.0886 (0.1427)
eco_proxy_4	0.3367 (0.1040) $***$	0.4204 (0.1590) $***$	0.2311 (0.1436)
eco_proxy_5	0.6697 (0.1282) $***$	0.6937 (0.2140) $***$	0.4856 (0.1679) $***$
eco_proxy_6	0.4685 (0.1468) $***$	0.5632 (0.2136) $***$	0.3419 (0.2102)
l_NB	-0.0119 (0.0162)	-0.0279 (0.224)	-0.0082 (0.0240)

Table 2: Number of Cases Correctly Predicted by Model 1

	Full sample	Born before 1982	Born after 1981
Number of cases	3920	2332	1658
Percentage	67.4%	71.1%	65.4%

Table 3: Distribution of Cases Correctly and Incorrectly Predicted by Model 1

Full sample

		Predicted	
		0	1
Actual	0	3184	444
	1	1453	736

Born before 1982

		Predicted	
		0	1
Actual	0	2216	99
	1	850	116

Born after 1981

		Predicted	
		0	1
Actual	0	910	403
	1	475	748