The Determinants of College Major Choice: Evidence from Iranian University Entrance Exam

Sepehr Ekbatani

University of California, Los Angeles

1 Introduction

There is a growing literature on the returns to education by major of study (Hastings et al. [2013],Kirkeboen et al. [2016], Altonji and Zimmerman [2017], Arcidiacono [2004]). However, we don't know much about the determinants of major choice and why students prefer one major over the others. Although several studies tried to answer this question by running surveys on college students about the characteristics of their favorite major, they lack a good data and also students are surveyed after they have chosen their major. (Arcidiacono et al. [2010], Wiswall and Zafar [2015], Baker et al. [2017])

These papers are based on questionnaires in which students are asked about their expected salary in future and comparing answers to actual labor market data to draw conclusions on lack of information and errors in expectations (Arcidiacono et al. [2010], Hastings et al. [2015], Wiswall and Zafar [2015], Baker et al. [2017],

Montmarquette et al. [2002]).

Another set of papers have found that non-pecuniary factors are the main determinants of major choice, while expected earnings play a significant but small role (Beffy et al. [2011], Wiswall and Zafar [2015], Stinebrickner and Stinebrickner [2013]). They find factors such as prestige, curriculum offerings, location, ... as the main non-pecuniary determinants of major choice. Because of data limitations, both set of papers have to look at broad major categories such as social sciences, engineering, etc. But the problem is that there is a lot of variance in both pecuniary and non-pecuniary factors among different engineering majors, for example.

By using a large dataset on ordered major choices and preference questionnaires, this study tries to fulfill the gap in the literature. Main feature of this dataset is that it includes the list of 100 ordered major/university choices for about 100'000 Iranian students. In addition to the list of ordered choices, students are asked about their main priorities for choosing a major and their expectations about future earnings of top majors.[1] Using a revealed preference approach important factors in major choice can be determined.

The article is organized as follows. In section 2, I will describe the secondary and post-secondary system of education in Iran and also the features of administrative and public data that I will use. Empirical strategy and results are discussed in section 3. Conclusion and further work are discussed in section 4. In Appendix A, some facts using the public data are shown and the other possible quesions that can be addressed are discussed.

[1]This administrative dataset is not public and needs to be accessed on-site.

2 Higher Education in Iran and Data

This section will provide a general description of secondary and post-secondary education in Iran. Figure 1 provides an overview of sequence of high school events. Details on the public and unique administrative data that is used in this study will also be discussed.

The problem with studies on major choice that are based on surveys and questionnaires is that results are usually influenced by the fact that the student has been already admitted to the school for following a specific major and it might influence his answers to major choice preferences. So, the ideal dataset for this project will include data on not only the major that student is admitted, but also the places and majors that he wanted to get into but he got rejected. There is a valuable dataset on students who take the university entrance exam (called Concours) in Iran which has all the properties needed to find the determinants of major choice which I will discuss in short.

At the end of the first year of high school, students have to choose among three broad majors, i.e. Mathematics and Physics, Experimental Sciences, Humanities and Literature. This will determine the set of courses that they will take in the following three years of high school. Mathematics and Physics students will have some exclusive courses such as Geometry, Calculus, ... Exclusive courses for Experimental Sciences include Biology, Geology, ... and for Humanities include Philosophy, Advanced Literature, ... Students in each broad major will be examined on general courses and on their own exclusive courses later on. This allows private high schools to specialize in one of the broad majors while public high schools are required by law to offer all three majors for their students which

results in a lower quality.

Students will have to participate in nationwide diploma exams at the end of third year of high school. Exams are held at the same time all around the country for everyone pursuing a diploma from one of the aforementioned broad majors. The scores on these exams will be a part of the final score for entering the university later on.

Fourth year of high school is not compulsory for those who don't want to pursue higher levels of education. They will receive their diploma by passing the diploma exams and can join the labor force. Those who plan to go to university have to sign up for the last year of high school, called pre-university year. Students will have regular school classes for the first six months of the educational year and have the other three for preparing for the important exam which is usually held in late June of each year.

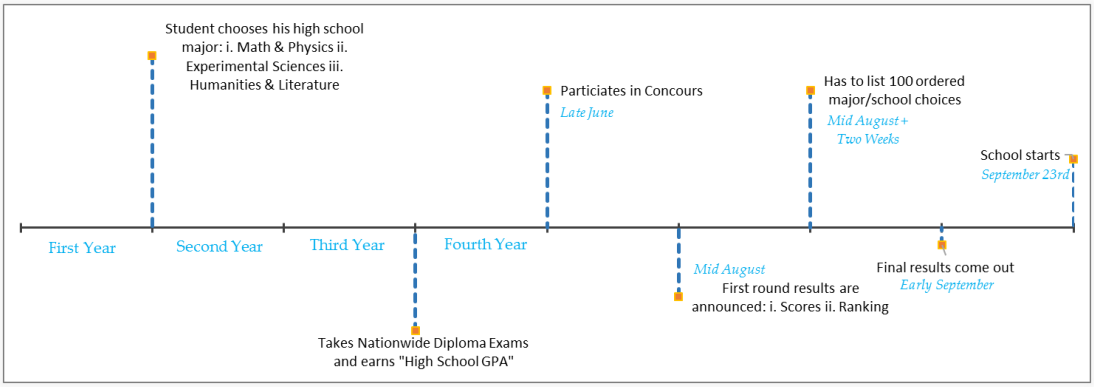


Figure 1: High School and Pre-University Timeline of Events

Concours is a 4-hour multiple choice exam in which for each broad major consists of the courses that students have taken throughout the four years of high school. Every year around 900'000 students (around 60% girls), who have com-

pleted four years of high school education participate in Concours. For entering university, students have to participate in Mathematics and Physics, Experimental Sciences or Humanities Concours based on their major in high school. Participation in Arts and/or Foreign Languages Concours is optional and all students can participate in them.

Pre-university is the most important grade for all students in Iran. Private school tuition costs for this grade exceeds an average worker's yearly income. In Tehran only, there are about 1800 test preparation institutions providing mainly afternoon classes, tutoring and mock exams. Revenue of these institutions is estimated to be around $2B (Almost 0.5% of Iran GDP).

As another evidence on importance of higher education, Table 1 compares the number of students per 100'000 in Iran with some similar countries in terms of GDP per capita along with United States and England.

	1975	1985	1995	2000	2005	2010
Iran	428	398	1557	2191	3103	5217
Turkey	690	870	2043	2548	3147	4147
Egypt	1053	1711	-	3779	3309	3337
Brazil	1039	-	-	1629	2503	3158
United States	4730	5158	5442	4735	5905	6673
England	1267	1807	3167	3478	3844	3969

Table 1: Number of College Students per 100'000
Source: UNESCO Education and Literacy Statistics

After taking the Concours, students have to wait for about 45 days for the first round of results to come out. Students will know their score on the exam which is a weighted average of their relative scores on different courses on the exam. Relative score means that student will receive a higher score if the average among

other students is very low. Based on these scores students will be ranked and the ranks will also be reported to the student. This ranking is the determinant of who gets to choose first his desired major of study.

After the announcement of scores and rankings, students will have two weeks to fill out a form with 100 choices of major/university (each major/university has a specific code, I call it course from now on) and submit it to the National Organization of Educational Testing (NOET). These choices are ordered and students would not benefit from listing any major other than their true preference. Most of students ask for consulting which will provide them access to the previous years results and the information on placement of students with similar rankings as theirs. So essentially they have a probability distribution of getting accepted to each major when they are making their choices. Since there is 100 rows to be filled they will have the option to choose their desired course even if they had no chance of getting in last year with the same score.

After this period, it is NOET's turn to assign students to the majors. System will start with the first person in ranking and will assign him to his desired choice. This process continues until the seats for a course is filled. From then on the students who has chosen that option as their first choice will be rejected and system will look at their second choice on the list. This process ends when either all the seats of all majors are filled or the last student in the ranking is reached.

The large number of choices that students can list make sure that the process is a one-sided Gale Shapely algorithm which is strategyproof. It proves that people will reveal their true preferences when they are making their choices. This is actually one the main differences between Iran's system with countries which are using similar centralized university entrance exams like Norway and Chile. In these

two countries number of choices that student can make is limited to 15 which makes it hard to ensure the strategyproofness. The other difference is that universities, both public and private don't decide on whom to accept and everything is done only through the NOET system.

The data on the number of seats for each major/university, rank and score of students who got accepted and their city of residence is public. The focus of this study would be on the administrative data which the list of 100 choices that students submit to the NOET in addition to many other individual characteristics. The data is classified and can only be accessed on site. Total summary statistics are shown in Table 2.

I have also used data from Household Income and Expenditure Survey (HIES) and Labor Force Survey (LFS) both conducted by Statistical Center of Iran to extract occupational earning and unemployment for years 2009-2015. Occupations and university majors are linked through International Standard Classification of Education 1997 (ISCED97) and Selected Characteristics of Occupations 1988 (SCO88). Table 7 shows all the ISCED major categories used in this study and examples of SCO occupations linked to them. Earnings and Unemployments are averaged with weights to form average salary and average unemployment of different categories. This is done for two reasons: i) Limited number of observations for some subfields in the data could have hurt the results of the paper. ii) Large number of categories make the estimation of the model very complicated and even impossible, so using these standards reduce the dimensions of the model and make it easier to solve.

Year	2009	2010	2011	2012	2013	2014	2015	2016
Number of Students	80'652	175'880	205'983	178'820	96'400	102'020	75'013	104'634
Number of Exam IDs	80'652	212'079	245'975	210'424	117'486	125'882	95'984	133'281
Average Number of Exams per Student	1	1.21	1.19	1.18	1.22	1.23	1.28	1.27
Average Number of Listings per Student	67	68	64	60	66	68	73	68
Total Number of Observations	5'401'227	14'367'653	15'663'283	12'576'816	7'719'203	8'604'346	7'010'853	9'048'905

Table 2: Total Summary Statistics

3 Empirical Strategy and Results

In order to find the determinants of major choice, I start with a simple OLS approach in which an index of the popularity of the major is produced and then regressed on associated labor market outcomes. I calculate the average ranking of students studying different majors and then regress this average ranking on labor market variables (Columns (1) and (2) of Table 3). This index might be affected by the capacity of majors in the way that if a major is only thought in the best school of the country then it would go first in the ranking and the average score of other majors would be affected by the students' scores of other schools and other cities. I define another index of popularity by counting number of top 10'000 students in the ranking who has chosen that major. The major with highest number of top students choosing it would be the highest in the ranking (Columns (3) and (4) of Table 3). The following equations show the naive model of major popularity:

$$Avg\ Rank_{jt} = \alpha_1 \log(Annual\ Salary)_{jt} + \alpha_2 Unemp_{jt} + \alpha_3 Growth_{jt} + \epsilon_{jt} \quad (1)$$

$$Top\ Students_{jt} = \beta_1 \log(Annual\ Salary)_{jt} + \beta_2 Unemp_{jt} + \beta_3 Growth_{jt} + \epsilon_{jt} \quad (2)$$

Table 3 shows that labor market variables such as income and unemployment rate affects the popularity of a major. Here I assume that students know about the labor market variables and they form their expectations based on the available

Table 3: Determinants of Major Popularity

	Average Rank of Students		Number of Students Ranked < 10000	
	(1)	(2)	(3)	(4)
Log Annual Salary	-5548.4	-7522.4*	47.47***	52.89***
	(3500.1)	(3718.8)	(14.04)	(15.09)
Unemployment Rate	1014.5***	1127.3***	-4.188***	-4.283***
	(260.4)	(268.4)	(1.045)	(1.089)
Salary Growth Rate		7144.6		-13.48
		(3921.9)		(15.91)
Constant	148424.1*	179834.2**	-682.2**	-776.0**
	(64274.2)	(67934.2)	(257.9)	(275.7)
Observations	571	549	571	549
R^2	0.029	0.040	0.044	0.046

Standard errors in parentheses

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

data. Columns (1) and (2) show that majors with higher income and lower unemployment rate are chosen by students higher in ranking. Same conclusion can be drawn by looking at the columns (3) and (4). The coefficients can be interpreted as if careers associated with major J earn one percent more annual salary than major J', J is on average chosen by about 50 more students in the top decile. Also, if the unemployment rate associated with major J is one percent lower than major J', J is on average chosen by four more students in the top decile. Both results show that majors in which currently are better in terms of labor market variables are chosen more by top students.

The OLS results show that labor market variables are important factors for a student who is deciding what major to study in the university. Based on this result we can use revealed preferences approach to estimate a model of major choice using the individual data. I assume the probability of choosing major j by individual i

to be:

$$Pr(y_i = j|x_i) = \frac{exp(x_i\beta_j)}{\sum_{k=1}^{J} exp(x_i\beta_k)} \tag{3}$$

By choosing $J = Humanities$ as the base category, logarithm of relative probabilities can be written as:

$$\eta_{ij} = \ln\left(\frac{Pr(y_i = j)}{Pr(y_i = J)}\right) = x_i\beta_j \tag{4}$$

Variables such as Age, Gender, High School GPA and City of residence are included in x_i. The trick used to identify the role of earnings in the probability of choosing a major is to use narrower categories for income compared to broader available choice categories. I assume that student chooses among ISCED97 codes which include 17 categories such as Humanities, Engineering, Law ..., while I use the major specific earnings based on SCO88 coding which includes 351 unique values for different majors. This allows me to have within group variation for earnings and to be able to add earnings to Equation 4. The Table 4 shows the estimation of Equation 4 using a sample of our data. The table is similar to Table 10 in Arcidiacono et al. [2010].

These result shows that girls are more likely to choose Humanities over Engineering, Architecture or Health. Students from small cities are less likely to choose Business and the probability that a high ability students chooses Engineering, Architecture and Health is higher than him choosing Humanities. It also shows that earnings affect the probability of choosing different majors. Although it is hard to interpret the coefficients, significant result shows that with change in the earnings, probability of choosing different major changes. This is another evidence for the

Table 4: Multinomial Logit Estimation of Major Choice

	ISCED97			
	Business & Addministration	Engineering	Architecture & Building	Health
Female	-0.209	-2.088***	-1.458**	-1.762**
	(-0.39)	(-3.78)	(-2.73)	(-2.89)
Mid Size Cities	-0.482	-0.124	-0.944	0.819
	(-0.84)	(-0.21)	(-1.64)	(1.22)
Small Cities & Villages	-1.336*	-0.614	-1.430*	1.202
	(-2.00)	(-0.89)	(-2.15)	(1.56)
High School GPA	0.118	0.405***	0.314***	1.443***
	(1.47)	(4.71)	(3.82)	(8.85)
Age	-0.164	0.00153	-0.0360	0.671*
	(-0.69)	(0.01)	(-0.15)	(2.31)
Log SCO88 Salary	8.785***	16.41***	9.957***	17.33***
	(8.35)	(13.37)	(9.08)	(13.67)
Constant	-160.9***	-311.9***	-187.2***	-362.6***
	(-7.90)	(-13.04)	(-8.80)	(-14.33)
Observations	728			

Humanities is the base category. Other categories are omitted to improve readability.
t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

importance of labor market outcome for major choice.

So far, we haven't used the main feature of the dataset which is the ordered preferences of majors. For this part, similar to Arcidiacono et al. [2010] and Zafar [2013] I will estimate a rank ordered logit model of major choice using the data in hand. Let $r_i = (r_{i1}, r_{i2}, ..., r_{i100})'$, where r_{im} is the major/university (course) that student i has ranked as the mth highest pair of 100 majors/universities. If v_{ijs} denotes the indirect utility function of student i from choosing major j in school s ($r_{im} = \{j, s\}$) and assuming that students choose their rankings to maximize their utility, probability of observing r_i for student i, would be:

$$Pr(r_i) = Pr(v_{ir_{i1}} > v_{ir_{i2}} ... > v_{ir_{i100}}) \tag{5}$$

Assuming that unobservable preferences follow a Type I extreme value distri-

bution, the probability that student i chooses ranking r_i is:

$$Pr(r_i) = \prod_{m=1}^{99} \frac{exp(v_{ir_{im}})}{\sum_{l=m}^{100} exp(v_{ir_{il}})} \tag{6}$$

And the log likelihood for the data would be:

$$L = \sum_{i=1}^{N} \log[Pr(r_i)] \tag{7}$$

Table 5 shows the maximum likelihood estimation of the rank ordered model by using 2015 data. Independent variable of the model would be the row number of the course in the student's list. To be more clear, higher ranked courses will have a smaller row number assigned to them. Main independent variables are SCO88 average salary and unemployment rate extracted from the 2015 LFS and 2015 HIES. I have also constructed a variable for studying in own state or another state. As column (1) of Table 5 shows, higher earnings and higher employment rate are associated with lower row number of the course in student's list. Also, students list the courses in their own state in rows with smaller numbers which means they prefer to stay in close to their hometown. Column (2) shows the fact that studying close to family is more important for girls compared to boys. The positive coefficient on *Female* variable states the fact that girls list the courses out of their state in rows with larger numbers. Column (3) shows that students from lower socioeconomic status are more willing to stay in their own state to study and larger coefficient on *Small Cities and Villages* shows its importance for poorest students. Note that better universities are mostly in large cities so column (3) might be an evidence of credit constrained students who cannot afford living far from their family.

	(1)	(2)	(3)
	Row	Row	Row
Log Annual Salary SCO88	-0.194***	-0.194***	-0.195***
	(0.00147)	(0.00147)	(0.00150)
Unemployment	0.0288***	0.0288***	0.0287***
	(0.000121)	(0.000121)	(0.000124)
Own State	-0.490***	-0.508***	-0.674***
	(0.00185)	(0.00274)	(0.00304)
Female		0.0389***	
		(0.00250)	
OwnState × Female		-0.435	
		(.)	
Mid Size Cities			0.340***
			(0.00298)
Small Cities & Villages			0.394***
			(0.00384)
Own State × Mid Size Cities			0
			(.)
Own State × Small Cities & Villages			-0.0406
			(.)
Observations	3713118	3713118	3585197

Standard errors in parentheses

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

Table 5: Rank Ordered Logit Estimation of Major Choice

Large number of observations cause the standard errors to be very low and the coefficient to be estimated with high precision. This is an important advantage of this study over other studies in the literature which are mostly based on surveys on a limited number of students. Another advantage is the school dimension of dataset which is rarely mentioned in the literature. As results show, students care about the labor market outcome and they also prefer to study close to their hometown. Most of the studies in the literature are done in the scale of one college or one university and they are unable to capture the importance of location of the school.

To make sure that the results are not sensitive to the major categories, I will reestimate the model using ISCED97 broad categories. The results are the same and are shown in Table 6.

	(1)	(2)	(3)
	Row	Row	Row
Log Annual Salary ISCED97	-0.147***	-0.147***	-0.145***
	(0.00264)	(0.00264)	(0.00266)
Unemployment	0.0343***	0.0343***	0.0342***
	(0.000121)	(0.000121)	(0.000123)
Own State	-0.469***	-0.492***	-0.644***
	(0.00180)	(0.00269)	(0.00296)
Female		0.450***	
		(0.00242)	
Own State × Female		-0.000221	
		(.)	
Mid Size Cities			0.323
			(.)
Small Cities & Villages			0.383***
			(0.00374)
OwnState × Mid Size Cities			-3.47e-09
			(0.00291)
OwnState × Small Cities & Villages			-0.0375
			(.)
Observations	3862037	3862037	3731132

Standard errors in parentheses

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

Table 6: Rank Ordered Logit Estimation of Major Choice Using ISCED97 Categories

4 Conclusion and Further Work

Using the unique dataset on major/university choices of Iranian students who took the nationwide university entrance exam, I estimated a rank ordered logit model of major choice. I showed that labor market variables, specifically earnings and unemployment play a significant role in choice of majors by students. Labor market variables were extracted from Labor Force Survey (LFS) and Household Income and Expenditure Survey (HIES) and were linked to different majors by International Standard Classification of Education 1997 (ISCED97) and Selected Characteristics of Occupations 1988 (SCO88). The model showed that students prefer majors with higher expected income and expected employment rate in the sense that they list these majors higher in their ordered ranking of majors.

Another finding of this paper is related to theories suggesting that many people might only care about the school and not the major. It can have several explanations, for example prestige of some schools might be one reason. Credit constraints of family or the cultural barriers might also play role for those students who prefer to stay in their hometown even at the price of studying a major that they are not very interested in. I provided evidence for these theories by including the dummy variable for studying in one's own state in my model. Significance of this variable and its interactions with gender and SES variables shows that there are such cultural barriers and credit constraints.

We are conducting a survey on students who took the Concurs in 2010 asking them about earnings and labor market status. This will give us the opportunity to run tests on how well people did on foreseeing future outcome and also let us run a regression discontinuity model to calculate returns to different majors.

References

Joseph G. Altonji and Seth D. Zimmerman. The Costs of and Net Returns to College Major. Technical Report 23029, January 2017. URL http://www.nber.org/papers/w23029. DOI: 10.3386/w23029.

Peter Arcidiacono. Ability sorting and the returns to college major. *Journal of Econometrics*, 121(1-2):343–375, 2004. ISSN 0304-4076. URL http://econpapers.repec.org/article/eeeeconom/v_3a121_3ay_3a2004_3ai_3a1-2_3ap_3a343-375.htm.

Peter Arcidiacono, V. Joseph Hotz, and Songman Kang. Modeling College Major Choices using Elicited Measures of Expectations and Counterfactuals. Technical Report 15729, February 2010. URL http://www.nber.org/papers/w15729. DOI: 10.3386/w15729.

Rachel Baker, Eric Bettinger, Brian Jacob, and Ioana Marinescu. The Effect of Labor Market Information on Community College Students' Major Choice. Technical Report 23333, April 2017. URL http://www.nber.org/papers/w23333. DOI: 10.3386/w23333.

Magali Beffy, Denis Fougère, and Arnaud Maurel. Choosing the Field of Study in Postsecondary Education: Do Expected Earnings Matter? *The Review of Economics and Statistics*, 94(1):334–347, May 2011. ISSN 0034-6535. doi: 10.1162/REST_a_00212. URL http://dx.doi.org/10.1162/REST_a_00212.

Justine S. Hastings, Christopher A. Neilson, and Seth D. Zimmerman. Are Some Degrees Worth More than Others? Evidence from college admission cutoffs in Chile. Technical Report 19241, July 2013. URL http://www.nber.org/papers/w19241. DOI: 10.3386/w19241.

Justine S. Hastings, Christopher A. Neilson, Anely Ramirez, and Seth D. Zimmerman. (Un)Informed College and Major Choice: Evidence from Linked Survey and Admin-

istrative Data. Technical Report 21330, July 2015. URL http://www.nber.org/papers/w21330. DOI: 10.3386/w21330.

Lars J. Kirkeboen, Edwin Leuven, and Magne Mogstad. Field of Study, Earnings, and Self-Selection. *The Quarterly Journal of Economics*, 131(3):1057–1111, August 2016. ISSN 0033-5533. doi: 10.1093/qje/qjw019. URL https://academic.oup.com/qje/article/131/3/1057/2461218/Field-of-Study-Earnings-and-Self-Selection.

Claude Montmarquette, Kathy Cannings, and Sophie Mahseredjian. How do young people choose college majors? *Economics of Education Review*, 21(6):543–556, December 2002. ISSN 0272-7757. doi: 10.1016/S0272-7757(01)00054-1. URL http://www.sciencedirect.com/science/article/pii/S0272775701000541.

Ralph Stinebrickner and Todd Stinebrickner. Academic Performance and College Dropout: Using Longitudinal Expectations Data to Estimate a Learning Model. Technical report, 2013. URL https://ideas.repec.org/p/uwo/hcuwoc/20135.html.

Matthew Wiswall and Basit Zafar. How Do College Students Respond to Public Information about Earnings? *Journal of Human Capital*, 9(2):117–169, June 2015. ISSN 1932-8575. doi: 10.1086/681542. URL http://www.journals.uchicago.edu/doi/abs/10.1086/681542.

Basit Zafar. College Major Choice and the Gender Gap. *Journal of Human Resources*, 48(3):545–595, July 2013. ISSN 0022-166X, 1548-8004. doi: 10.3368/jhr.48.3.545. URL http://jhr.uwpress.org/content/48/3/545.

Appendix

A Descriptive Results

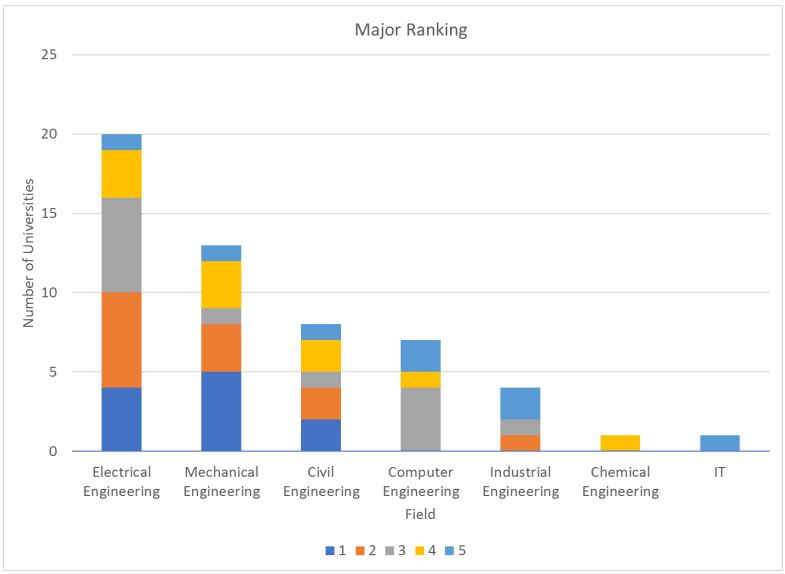
For finding the most popular majors, for each university, I have sorted the ranking of last person who has filled the seats of that major in that specific university. This has been done for all 12 universities for 2013 and 2015 Concours and the results are shown in Figure 2.

As shown in Figure 2a, Electrical Engineering was by far the most popular major in 2013. Different fields of electrical engineering (Electronics, Communication, Power) were the first major to be filled in 4 universities, filled second in 6 and filled third in 6 universities. Mechanical Engineering, Civil Engineering and Computer Engineering were the next popular majors in 2013.

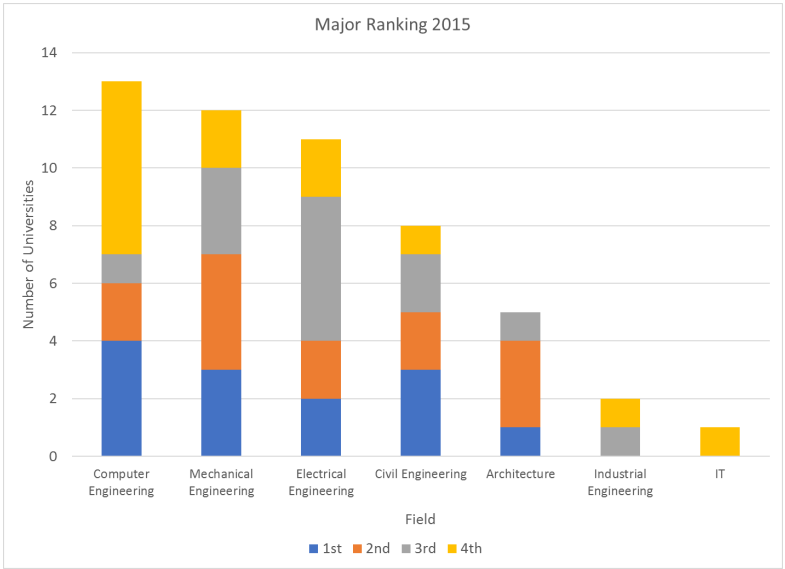
Comparing Figure 2a and Figure 2b shows that 4 most popular majors are the same as 2013 in 2015 but the rankings has changed. Electrical Engineering which was the most popular major in 2013 by far, became third in ranking in 2015. On the other hand Computer Engineering which was the fourth popular major in 2013, became the most popular one two years later. It seems that students have responded to the growing labor market demand for computer engineers.

The interesting fact is that similar pattern can be seen among the UCLA applicants too. Figure 3 shows the share of transfer applicants to different UCLA engineering majors for the past 15 years. As can be seen, Computer Science has became very popular in recent years, while Electrical Engineering and Civil Engineering are not very popular anymore.

Intuitively, it seems that students respond to the signals from the labor market



(a) 2013



(b) 2015

Figure 2: Majors Ranking

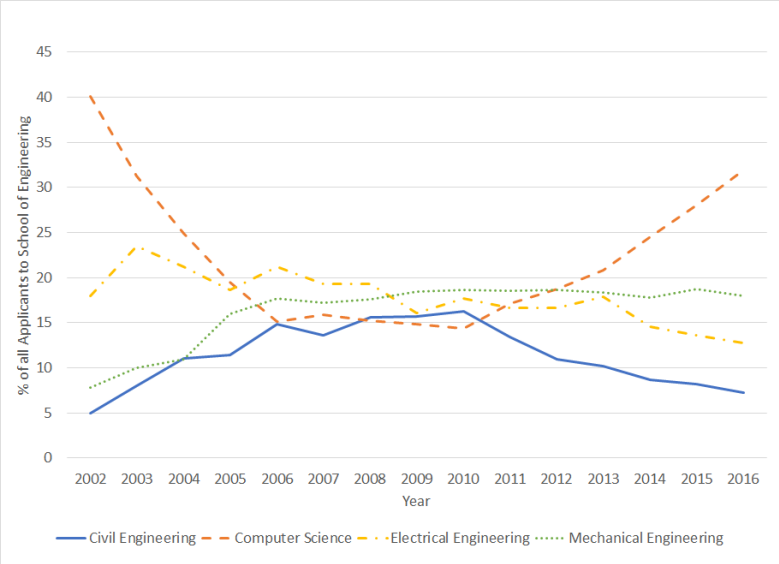


Figure 3: Evolution of Share of All Applicants to Engineering School
Source: UCLA Undergraduate Admission Profile of Transfer Applicants

and those majors which related to growing occupations become more popular among the students. It is important to see whether this major ranking is the same for average students too and whether average students respond to labor market trends similar to top students or not. This raises the question that to what extent people care about the labor market outcome when they are choosing their major? And also how responsive major choice is to labor market changes? These questions cannot be answered with this limited public dataset and needs access to the administrative one.

If salaries are added to Figure 2a we can see that relationship between the most popular major and most paid occupations is not very clear. Using the Consumer Expenditure and Income Survey and averaging the income of workers who graduated with a degree in one of 18 majors, salaries associated to each majors' graduates has been found. Figure 4 shows ranking of majors with the associated

salary. It can be seen that although some majors are very popular among students but the graduates are not among the top paid workers. For example, Mechanical Engineering and Civil Engineering are second and third in the majors ranking, but the graduates are paid less than Computer or Chemical Engineering graduates, which are not among the top majors.

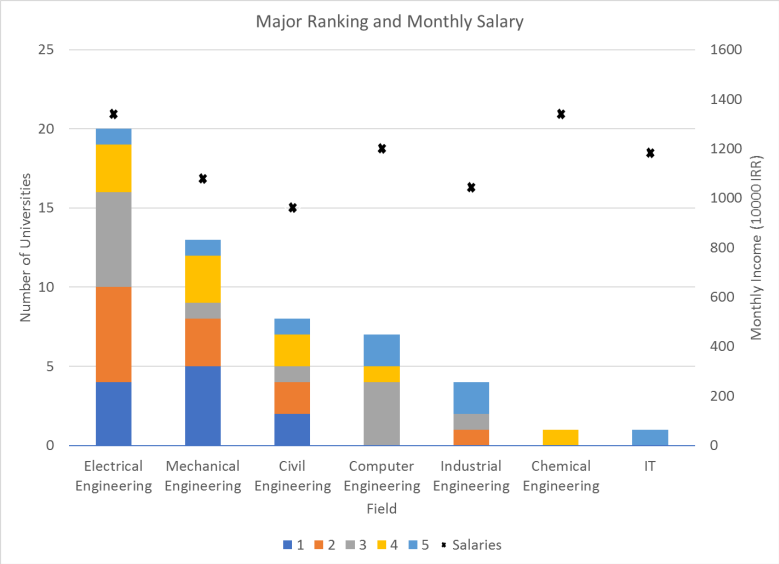


Figure 4: Majors Ranking and Monthly Salary 2013

Next figure provides evidence for what is referred to as "school prestige" in the major choice literature. For each major, universities have been sorted based on the ranking of last person who got a seat in that university. As can be seen in Figure 5, Sharif University is the most prestigious school in Iran. All majors that are offered in Sharif are filled first among all the universities. Another important observation is that top five universities are all in Tehran which suggests that quality of schools might be very unequally distributed in the country.

Figure 6 shows the share of non-resident students in each school in the sample. This is an important evidence suggesting that people might choose schools rather

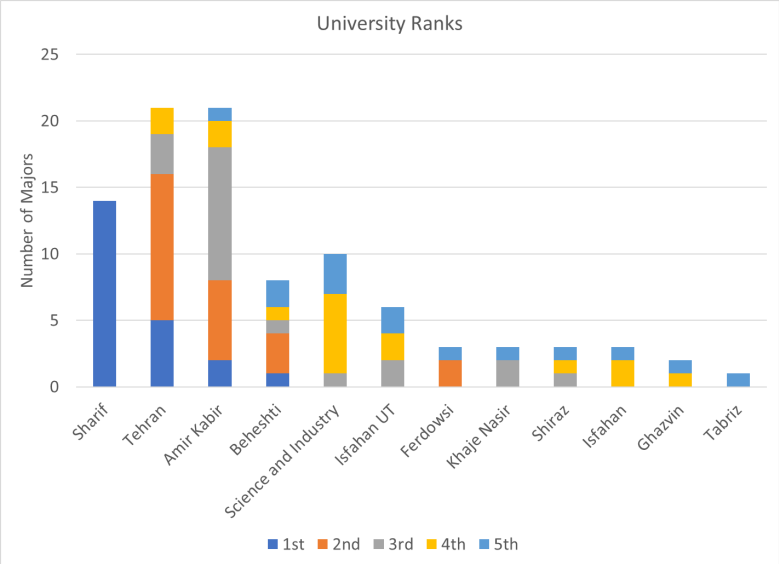


Figure 5: School Ranking

than majors. Share of non-resident students in universities in Tehran shows that students from other cities are willing to leave their home town for the capital only if they got accepted to Sharif University. However, this is not true about the universities in cities other than Tehran. The share of non-residents in those universities are relatively high with most of non-residents coming from Tehran. This suggests that families from cities other than Tehran might face some kind of a credit constraints that the students would prefer to stay at their hometown.

Figure 5 and Figure 6 put emphasis on an important question in the literature: Do people choose major or they choose the school? This question can be answered very clearly using the data on 100 ordered choices of students. To be more clear on this, if one student has listed one major in different universities as her top choices, it shows that she gives priority to the major. On the other hand, if one student has listed all majors in one university at the top of his list, he is definitely choosing school and not the major. The difference here might be very important because

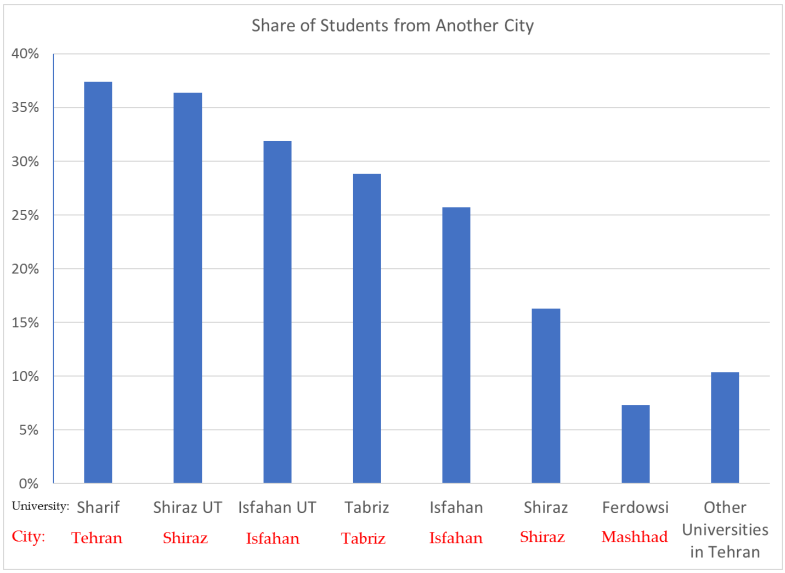


Figure 6: Share of Non-Resident Students in Different Schools

in the literature it is always assumed that the major that student is studying in college is his first choice and other majors are his alternatives.

This hypothesis should also be tested that, does the share of students who choose major and the students who choose the school differs among top students and average students. It seems possible that top students who can choose among more options, choose their desired major but average students choose the school and not the major.

In this section, public data on Mathematics and Physics Concours were used to provide some descriptive results which can be suggestive for what interesting questions to answer using the administrative dataset.

B Standard Classifications

Category (ISCED97)	Major Example (SCO88)
Teacher training and education science	Teacher Training
Arts	Cinema
Humanities	Foreign Languages
Social and Behavioural Science	Economics
Journalism and Information	Media
Business and Administration	Accounting
Law	Law
Life Sciences	Biology
Physical Sciences	Geomorphology
Mathematics and Statistics	Statistics
Computing	Computer Science
Engineering and Engineering Trades	Electrical Engineering
Manufacturing and Processing	Mining Engineering
Architecture and Building	Civil Engineering
Agriculture, Forestry and Fishery	Agriculture
Veterinary	Veterinary
Health	Medicine

Table 7: List of Categories and Major Examples