Essays in College Admissions and College Major Choice

By

Hayri Alper Arslan

Dissertation
Submitted to the Faculty of the
Graduate School of Vanderbilt University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY
in
Economics

June 30, 2018

Nashville, Tennessee

Approved:

Tong Li, Ph.D.

Eric W. Bond, Ph.D.

Andrew Dustan, Ph.D.

Eun Jeong Heo, Ph.D.

Mümin Kurtuluş, Ph.D.

To my family and my beautiful country, Turkey

ACKNOWLEDGMENTS

I thank the Department of Economics at Vanderbilt University for their generous financial and educational support over the past 6 years. Apart from financial support, I am thankful for very welcoming environment that let me being part of this organization. I am also grateful for financial support from Tong Li, the Noel Dissertation Fellowship, and Kirk Dornbush Summer Research Grant.

There are many people who have been very important in supporting me academically and personally while I have been in graduate school. First, I am particularly grateful to members of my dissertation committee. Professors Eric Bond and Andrew Dustan have been accessible and provided me with much valuable feedback on my work. Their feedback have been improved my work a lot. Professor Eun Jeong Heo provided excellent insights about college admissions models and many problems that I am addressing in my dissertation are related to her feedbacks. I am also grateful for her accessibility, understanding, and friendship. I have also benefitted from being the student of many great teachers along the way: Ahmet Alkan and Firat Inceoglu. They are the two people who provided invaluable college admissions data and made me start to work on empirical college admissions.

Second, I am extremely fortunate to be one of the students of my advisor, Tong Li. He does not only guide me in this long journey but also provide me an excellent environment to become an independent researcher in topics, which I am interested. His clear mind in any topics and his suggestions have always helped me to find a better way to think, even though I have realized many of them after a long time. I am grateful for his patience, generosity, and support.

Third, I would like to thank my friends at Vanderbilt University who provided great help on my numerous things and more importantly make my experience at Vanderbilt more enjoyable.

Finally, I am thankful to my Mom and Dad. They have always supported and helped me. Even though there was long distance between us, they give me power to work harder. I especially thank my brother, Muzaffer. We have been going over similar steps in the career our paths from the beginning of our lives. His friendship, support, encouragement is very important for me. I am grateful for his reading of my stuff over and over again, providing a lot of great feedback.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Choosing college and major is important decision as it shapes individuals' earnings, working conditions, and lifestyles. Economists are interested in knowing determinants of college admissions and how various mechanisms and admissions policies affect students behaviors for a long time because understanding underlying factors, motivations, and strategies are crucial to design better mechanisms and education policies. The aim of this dissertation is understanding students' choice behavior in college admissions and major choices by providing new empirical methods. In this line, I develop empirical methods to understand students' college preferences from their applications, the effects of their preferences on their college admissions preparation strategies, and to tests the effects of students' marriage expectations on their college major choices.

In my first chapter, I propose a practical and data-driven approach for estimating students' college preferences from reported lists in centralized admissions mechanisms based on the Deferred Acceptance Algorithm. The Deferred Acceptance (DA) is one of the popular admissions mechanisms, however, it is applied with restricted rank-order lists (ROL) in many real world examples, which is different than its advocated version. Under restriction, reporting truthfully is no longer one of the best strategies. Moreover, there are many evidences of strategic behavior in students' college rankings even among the students that do not use all listing in their ROLs. In order to incorporate strategic behavior in the estimation strategy, I model students' ROL preparation strategies. Combining the prediction of this model with the observed stylized facts in the data, I develop an estimation method based on students' predicted ranking behavior where students are making their lists within the sets of colleges that are expected to be considered and found to be accessible. This setting enables to make comparisons by excluding colleges that are eliminated from students' lists because of their assignment beliefs. In addition, I can estimate students' college preferences using complete information in the reported ROLs. Finally, with the predicted preference orders, I document the magnitude of welfare changes of using long-term versus short-term student performance measures by income categories, high school types, gender, and age groups.

The second chapter analyses the effects of students' colleges preferences on their college admissions preparation strategies and provides empirical methods to test its significance and magnitude in the data. I focus on centralized college admissions, where colleges sort students according to their scores from a national entrance exam. The outcomes of the college admissions with entrance exams can be represented with cutoff scores, which are the minimum scores of the matched students. The distributions of cutoff scores are good approximations for matching outcomes in large matching games rather than complicated models and students' behavior can be characterized in individual level using these distributions. In this environment, I show that students' exam preparation strategies depend on their abilities, personal college valuations, and the distribution of cutoff scores, where students approximate it using cutoff scores from the past years. This setting allows me to examine the effects of students' college preferences on their exam preparation strategies in the data. To identify the effects of students' preference on the preparation strategies and resulting exam scores, I partition colleges according to observable college characteristics and competition levels. These partitions also enable to estimate the effects of preferences on exam scores. Using Turkey's college admissions data, I show that students' college preferences are significant factors in the exam preparation strategies and they account for up to 40 percent of the total variation in exam scores. This finding suggests that disregarding the effects of students' college preferences on the exam preparation strategies causes misleading interpretations of students' behavior in college admissions. In addition, admission policies (e.g., Affirmative Action and Financial Aid) need to be reconsidered according to the effects of preferences.

In the third chapter, co-authored with Tong Li, we test the effects of marriage expectations on college major choices. The data from National Longitudinal Survey of Youth 1997 (NLYS97) show that there are significant differences in marriage realizations of college graduates from different majors. Selection of a major with different marriage expectations can present one potential explanation for why there are observed differences in the marriage outcomes. To test this hypothesis, we develop a copula-based econometric framework that incorporates multinomial regressors (major choice) in binary outcome response (marriage realization) models. Our test results show that the effects of marriage expectations can not be rejected, even after individual characteristics and expected earnings are controlled for.

These three chapters provide new methods for understanding students' behavior in college admissions and empirical evidence of students' preferences, and motivations' effects on their application strategies, and their preparation strategies. These results show that students with different preferences and motivations prepare and choose education paths accordingly. Moreover, admissions criteria, admissions policies (e.g., Affirmative Action and Financial Aid) affect the welfare of students from different backgrounds differently and they need to be reconsidered according to policy makers' objectives.

CHAPTER 1

Preference Estimation in Centralized College Admissions from Reported Lists

1.1 Introduction

Understanding students' college preferences is crucial for evaluating performance of admission mechanisms and measuring the effects of admission policies on students' welfare. In this line, centralized college admissions provide rich administrative datasets for estimating students' college preferences. In these mechanisms, students and colleges are matched according to students' reported rank-order lists (ROL) and admissions criteria (e.g., exam score, high school GPA) that sort students. Even though there is rich information about students' college choices coming from reported lists, recovering preferences from these lists is not immediate. Reporting college lists truthfully is not one of the best strategies in many real world admissions mechanisms, and therefore estimating preference by treating the colleges in the submitted lists as if they are the most preferred colleges is not supported by theoretical findings.[1]

In this paper, we propose a practical and data-driven preference estimation method from reported ROLs in a centralized college admissions based on the Deferred Acceptance (DA) without assuming truthful revelation of preferences. The Deferred Acceptance is one of the popular admission mechanisms, however, it is applied with restricted ROLs in many real world examples, which is different than its advocated setting.[2] If there are restrictions on the number of reported colleges, reporting truthfully is no longer one of the best strategies (Haeringer and Klijn (2009), Calsamiglia et al. (2010)). Moreover, there are many evidences of strategic behavior in students' college rankings even among the students that do not complete their ROLs.[3] Therefore, in order

[1]In addition, we observe that students omit popular colleges if they expect a low chance of being accepted in our data. Omission of popular colleges is a supporting evidence that truth-telling is not a good assumption to rely on for estimating preferences.

[2]Mechanism design literature examines centralized admissions mechanisms extensively (Sönmez and Ünver (2011), Abdulkadiroglu (2013)) and advocates mechanisms that incentivize truthfully reported lists (Pathak and Sönmez (2008)). The DA is one of the advocated mechanisms because reporting rank-order lists (ROL) truthfully is the one of the best strategies if students are allowed to rank all colleges.

[3]Majority of students do not report complete lists in college admissions in Turkey, see figure A.5. Students' scores

to incorporate expected strategic college ranking behavior in our estimation strategy, we start by modeling students' college ranking strategies in ROLs. Combining the prediction of our model with the observed stylized facts in the data, we develop an estimation method according to personal choice sets where students are making their lists among the colleges that are expected to be considered and found to be accessible.[4] This setting enables us to make comparisons by excluding colleges that are eliminated from students' lists because of their assignment probability beliefs. In addition, it allows us to estimate students' college preferences using complete information in the reported ROLs.

We use a large college admissions model to analyze students' college ranking strategies. Large admissions models allow us to characterize students' optimal college ranking strategies in individual level according to expected matching outcomes. In many national level admissions, matching outcomes, which are cutoff scores of colleges in score-based admissions, are publicly available and highly correlated throughout years. Since expected outcomes can be predicted from the past years' outcomes, adopting a model that incorporates public information makes our method data-driven compare to other approaches that depend on theoretical results. Moreover, this characterization is a better representation of students' ROL preparation behavior in national level college admissions because students are also benefiting from the past years' cutoff scores.

Implementing the proposed students' college preference estimation method requires to construct a personal consideration set for each student. In centralized college admissions, students may not include all their preferred colleges because they may believe that they have no chances for some of those colleges.[5] Therefore, there is high chance of reaching misleading conclusions by treating all colleges that are not listed in the ROLs as if they are not preferred by students. This is the major concern for estimating students' college preferences by adopting truth-telling assumption. Our method, however, is designed to exclude the colleges that are not included in ROLs

---

and listed cutoff scores of colleges are highly correlated, see the table A.11

[4]An analogous approach is developed using browse sets of users in the estimation of mate preferences in online dating by Hitsch et al. (2010).

[5]These cases can be also occurred in a centralized admissions with unrestricted lists because of students may not include colleges if students have information about assignment probabilities and they believe they have no chances to be assigned.

because of students' assignment probability beliefs. In this setting, all colleges in the consideration sets could be selected and ranked in students' ROLs, however, students rank them according to their true college preference orders. As a result, the proposed method do not suffer from inaccurate comparisons in the estimation.

There are equilibrium-based estimation methods also proposed to overcome problems resulting from truthful revelation of preferences assumption (Among others, Fack et al. (2017), Akyol and Krishna (2017), Burgess et al. (2015)). Stability/justified-envy-free is an well accepted equilibrium notion in the two-sided matching games and its implications are used in students' preference estimation methods.[6] Even though these methods provide elegant solutions for estimating students' preferences, robustness of stability-based estimation depends on whether students play equilibrium strategies or not. If students do not play equilibrium strategies, the matching outcome is less likely to be stable. Estimating students' college preferences based on stability assumption is rather restrictive because this approach does not incorporate individual mistakes and unobservable choice heterogeneities (e.g., understanding of admission mechanism) among students. In addition, resulting estimation methods use only matched colleges, which causes loss of valuable information in the reported lists. Finally, it is observed that around 20% of students who submit lists do not matched with any colleges. Estimating preferences based on stability has to exclude these students, which causes another under utilization of available data.

The proposed estimation method extends and improves stability-based estimation frameworks by using information from past years' matching outcomes. Outcomes from large matching markets include valuable information if agents benefit from this information when they are planning their strategies. Specifically for national level admissions, it is observed that past years' matching outcomes are stable throughout years.[7] In this setting, students' college preferences and exam scores are private information, however, past years' outcomes are generally publicly available and provide

[6] Stability implies that students are assigned to their favorite feasible colleges and a college is feasible in a exam-based centralized college admissions for a student if the minimum score of all admitted students (cutoff score) of the college is lower than the student's score.

[7] Data from Turkish college admissions shows that the stability of cutoff scores throughout the years. Table A.18 presents the correlation of cutoff scores throughout years and figure A.6 shows the distribution of demeaned cutoff scores. Ajayi (2015) also documents that the stability of matching outcomes in Ghana.

information for the forthcoming matching outcomes. Under these conditions, it is expected that candidates prepare their ROLs using available information from the realized matchings. We use past years' matching outcomes as the main information source to construct personal consideration sets, which are the crucial part of our estimation strategy. In addition, we personalize constructed consideration sets by allowing individual heterogeneities (e.g., understanding of mechanism) in ranking behavior. Therefore, the proposed estimation method not only relaxes strong symmetric equilibrium-play assumptions but also provides a framework to include complete information in the reported ROLs.

We apply our estimation strategy to centralized college admissions data from Turkey. In Turkey, the national admissions center assigns students to colleges based on students' reported ROLs, college entrance exam scores, and high school GPAs using DA. The number of available college-majors slots are more than 7500.[8] Students are allowed to rank up to 24 colleges. Our data-set consists of 12829 students, however, for the estimation we restrict analysis with students who have reported their rank order lists after learning their scores.[9] This data-set is a combination of a questionnaire that includes students socioeconomic background information and their performances in the college entrance exam and high school. We also collect cutoff scores of colleges from the national admissions center website in order to benefit from public information.

To compare our consideration set based estimation method with other methods in the literature, we use two methods. First one is the Hausman-type specification test. As in Fack et al. (2017), truth-telling vs stability comparison test, we implement a test for stability assumption vs consideration set approach. Our test rejects truth-telling and stability in a favor of consideration set approach. The second method is comparing predictive power of the estimators. We use college-major slots from the private colleges that requires different scholarships, however, provide the same education in order to create prediction power measure. The comparisons of estimated rankings show that consideration set approach has higher predictive power such that it has better performance in the

[8]Major and college choice are simultaneous.

[9]We exclude students, who have two year programs in their ROLs. Two-year programs have special admissions rules and it is hard generate a uniqued framework for their priorities.

prediction of students' college-major preference orders. Finally, we can summarize estimation results according to our preferred estimation results. The results show that colleges in the same cities and private colleges that provide scholarship are more preferred. College programs that are offered in evening hours are less preferred. Moreover, these preferences are changing with gender, income levels, and parents' education levels.

Finally, preference estimates provide an opportunity to perform counterfactual analysis. There are many rules for admissions criterion in college and public school admissions. Since these criteria are one of the main policy instruments for the admissions mechanisms, it is important to see their welfare effects. By predicted complete preference orders, we evaluate only college exam score-based versus only high school GPA-based admissions criteria's effects on students' welfare. We find that the magnitude of welfare changes by income categories, high school types, gender, and age groups. These results suggest that policymakers should take differentiated effects of admissions criteria when they are designing college admissions.

The structure of this paper as follows: We conclude this section with a discussion of the related literature. In section 1.2, we describe a centralized college admissions mechanism. In the 1.3 section, we present the proposed estimation strategy that is based on personal consideration sets. Next, we describe our econometric specification. Section 1.5 presents the college admissions mechanism in Turkey, data, and estimation results. In section 1.6, the counterfactual analysis of different admissions criteria is presented. Then, we conclude.

1.1.1 **Related literature**

From the influential papers by Balinski and Sönmez (1999), Abdulkadiroglu and Sönmez (2003), the design of college/school admissions mechanisms has been an important research path, and many admissions mechanisms are changed after these and following studies. Even though in some special cases depending on truthful revelation assumption is supported from data and the mechanism (Drewes and Michael (2006), Abdulkadiroğlu et al. (2017)), it is not applicable in most of the real world examples.

Our paper not only develops an estimation method based on large matching models such that it utilizes all choices in ROLs and all students who report lists but also relaxes dependence of equilibrium play assumptions by allowing individual heterogeneities and benefiting from available public information. Even though Fack et al. (2017) deals with truth-telling assumption rigorously by estimating preferences based on asymptotic stability, they only utilize matching outcomes and feasible sets[10] of colleges to estimate students' preferences. Kirkebøen et al. (2012) and Burgess et al. (2015) also use the same discrete choice framework to understand the effects of educational choices. These approaches do not utilize all available information in the reported ROLs. However, we are incorporating all listed colleges in ROLs by extending estimating framework based on students' behavior in large college admissions model and constructed personal consideration sets.

Among other papers that are proposing estimation of students' preferences from reported ROLs, Ajayi (2015) is the closest one to our paper. She also creates sets of schools to estimate preferences according to public information and selectivity criteria. Different than the proposed estimation framework in this paper, she adopts stronger choice assumptions and set restrictions in the construction of choice sets.[11] Our consideration set-based estimation, however, are supported by large college admissions model's predictions.

Large matching markets, which are used to characterize students' college ranking behavior, is another related literature. Azevedo and Leshno (2016) develop a supply and demand framework to analyze stable matchings by assuming a continuum of students. The continuum version of the model typically has a unique matching and is a good approximation for real markets. Menzel (2015b) works on two-sided matching markets with non-transferable utility when the number of market participants grows large and show identification conditions in the large markets. Azevedo and Budish (2017) examines strategy-proofness in the large, and argue that it is a useful second-best to exact strategy-proofness for market design. The large market framework is also studied in other game theoretical models to characterize agents' behavior. Bodoh-Creed and Hickman (2016)

[10] All colleges that have lower cutoff than a students' priority are defines as feasible colleges.

[11] These set restrictions are not supported in our data and ranking of schools according to assignment probabilities is not supported by data also it does not represent underlying valuation of schools.

analyze affirmative action policies by modeling college admissions as a large contest. Menzel (2015a) proposes inference framework, the limiting distribution of players' choices and characteristics is equivalent to a single-agent discrete choice problem. These papers are the recent examples of using large market framework to overcome complicated analysis in choice and game-theoretical settings.

Apart from preference estimation in the mechanisms using DA, there is also growing literature on students' preference estimation under different admissions mechanisms. Among others, He (2012), Calsamiglia et al. (2014), Hwang (2015), and Agarwal and Somaini (2016) examines agents' behavior in admissions using Boston Mechanism. Since agents can benefit from changing their true preference orders in Boston Mechanism (Immediate Acceptance), estimating their preferences also requires to understand underlying students' college/school ranking behavior. In these papers, reported ROLs are considered the one that maximizes expected utility of agents. Then by using calculated assignment probabilities, they apply discrete choice methods. Different than these papers, we propose our estimation method by constructing personal choice sets based on subjective assignment probabilities.

Researches with data from decentralized college admissions are another related literature. Avery et al. (2012) propose a ranking of U.S. undergraduate programs based on students' revealed preferences. Long (2004) examines how students choose colleges in U.S. and answer decisions determinants changes over time.

Theoretical studies on choice and considerations sets are also related to our paper. Manzini and Mariotti (2012) propose a boundedly rational model of choice where agents categorize alternatives before choosing. Kimya (2018) analyzes choices of decision makers that use relative ranking of the alternatives on each attribute to reduce the size of the choice set. The same reduction principle can also be thought in college choice case, where students face many options and they filter their choice sets using expected college cutoff scores. Similarly, Demuynck and Seel (2018) derives revealed preference tests for models where individuals use consideration sets to simplify their consumption problem. Different than these papers, our paper characterizes students' college ranking problem within their consideration sets and propose an estimation methodology.

Finally, college admissions data from Turkey are also studied in other related papers. Krishna et al. (2017) study central college admissions and analyze the effects of retaking exams. They develop and estimate a structural model of exam retaking using data from Turkey's college admissions. Caner and Okten (2013) examine how the benefits of publicly financed higher education in Turkey are distributed among students with different socioeconomic status. Saygin (2013) finds that gender has significant effects on college choice in Turkey such that female students prefer lower ranked colleges. However, none of these papers investigates the estimation methodology under DA with truncated ROLs.

1.2 Model

College admission is a classical example of matching markets and matching theory provides a structure to analyze students' college ranking behavior. In this section, we start by summarizing a centralized college admission setup. We first analyze students' college ranking behavior when the number of students are finite, and then show convergence to the limit game. The limit game is defined with the continuum of students and with the same proportions of college capacities as in the finite game. Students' college ranking strategies are analyzed using the result from the limit game approximations and we also assume that students use past years' outcomes as the main source of information when they are forming their expectation for the forthcoming matchings.

1.2.1 Centralized college admissions

We consider a student-college matching with a set of students (S), which is indexed by $i \in \{1,\ldots,n\}$, and set of colleges (C), which is indexed by $j \in \{0,\ldots,J\}$. College 0 denotes being unmatched. Each college j has $q_j > 0$ seats and these are announced before the students' exam scores[12] are known. C_0 has unlimited seats. We assume that student i derives $u_{i,j}$ utility from

[12] In college admissions model where colleges are passive in the matching, colleges' student preferences are formed from admissions criterion and they are called priorities. We normalized priorities using the $[0,1]$ interval.

assignment into college j. Utility of being matched with each college is assumed to have support $[\underline{u}_j, \overline{u}_j] \in [0,1]$ for all $j \in C$. Colleges rank students according to entrance exam scores, which are transformed into priorities $p_{i,j}$. A college j prefers student i over student k if and only if $p_{i,j} > p_{k,j}$.

After announcement of entrance exam scores, each student has an option to submit a rank order list (ROL) of $L_i = (l_i^1, \ldots, l_i^K)$. K denotes the length of the ROL and it is smaller than the number of colleges in the mechanism ($K < J$). l_i^k is student i's k^{th} choice and it is assumed that $u_{l_i^a} > u_{l_i^b}$ if $a < b$. Finally, we assume that there is a positive cost of including a college in ROL and denoted with $\kappa > 0$.[13]

The centralized student-college matching is solved by a mechanism that takes students' ROLs and exam scores that determine students' priorities. In this paper, we focus on a student-proposing Deferred Acceptance algorithm that is aligned with our data. The algorithm works as follows:

- *Round 1*: Every student applies to her first choice. Each college rejects the lowest-ranked students in excess of its capacity and temporarily holds the other students.
- *Round N*: Every student who is rejected in *Round N-1* applies to the next choice on her list. Each college, pooling together new students and those who were held from *Round N-1*, rejects the lowest-ranked students in excess of its capacity.

The process terminates after any Round N when no rejections are issued. Then each college is matched with the students that it is currently holding. A matching is defined as in Balinski and Sönmez (1999) such that it is an allocation of colleges to students such that no student occupies more than one position. Formally;

Definition 1.1 *A **matching** is a function $\mu : S \to C$ such that $|\mu^{-1}(C_j)| \leq q_j$ for all $j \in C$. If $\mu(i) = C_0$, student i is not assigned any college.*

If a student is rejected in all rounds of the algorithm, she is assigned to C_0 in the model, which means she is not assigned to any college.

[13]The cost of adding a college can be considered as a search cost. Students' search for colleges and majors is not only choosing form available set that is provided from placement center but also it is inquiry for college qualities, facilities, and major related opportunities and projections.

1.2.2 Students' college ranking strategies in the finite economy

In student-college matching, students have assignment probabilities to be matched with each college. These probabilities depend on their priorities and reported ROLs as well as their rivals' priorities and ROLs. If one expresses students' assignment probabilities according to priorities and ROLs, students' behavior can be analyzed according to expected utilities. Following Fack et al. (2017), Agarwal and Somaini (2016), assignment probabilities are expressed by;

$$\Omega_{i,j}(L_i, p_i; L_{-i}, p_{-i}) \equiv \begin{cases} Pr(\text{i is rejected by } l_i^1, \ldots, l_i^{k-1} \text{ and accepted by } l_i^k = j | L_i, p_i; L_{-i}, p_{-i}) & \text{if } j \in L_i \\ 0 & \text{if } j \notin L_i \end{cases}$$

Given the matching algorithm, $\Omega_{i,j}(L_i, p_i; L_{-i}, p_{-i})$ is either one for some college in C or zero if there is no ties in priorities.[14] Since priorities and preferences are private information, students form their strategies according to joint preference and priority distribution $G(u|p) \times H(p)$.

Student i's strategy for ROL is a function of her utilities and priority, $\sigma(u_i, p_i) : [0,1]^J \times [0,1] \rightarrow \Delta(L)$. In the symmetric equilibrium, σ^* solves the following problem for every student:

$$\sigma^*(u_i, p_i) \in \underset{\sigma}{arg\,max} \left\{ \sum_{j=1}^{K_i} u_{i,j} \int \int \Omega_{i,j}(\sigma, p_i; \sigma^*(u_{-i}, p_{-i}), p_{-i}) dG(u_{-i}|p_{-i}) dH(p_{-i}) - \kappa K_i \right\} \quad (1.1)$$

Existence of pure and mixed strategy Bayesian Nash Equilibrium are established by Milgrom and Weber (1985), while uniqueness is not generally true.[15]

Given an equilibrium σ^*, a matching μ_{σ^*} is observed such that the outcome of matching is summarized with the ex-post cutoff of each college:

[14] We assume strict priority orders based on exam scores. In priority rules with ties, characterizing assignment probabilities requires special format, which is not considered in this paper.

[15] Multiple equilibria and multiplicity of stable outcomes also increase the chance of students play non equilibrium strategies or play different equilibria strategies. Consequently, the chance of stable outcome decreases.

$$t_j(\mu_{\sigma^*}) = \begin{cases} min\,\{p_{i,j} | i \in \mu(j)\} & \text{if } \; j \notin \mu(j) \\ 0 & \text{if } \; j \in \mu(j) \end{cases} \tag{1.2}$$

where $t_j(\mu_{\sigma^*})$ denotes the cutoff of college j. Equation 2.2 shows that if there is no remaining slot after the matching, cutoff is the minimum priority level of the matched students. Zero cutoff means that particular college does not receive sufficient application to meet its capacity. Then, assignment probabilities can be redefined with the cutoffs such that;

$$\Omega_{i,j}(L_i, p_i; L_{-i}, p_{-i}) = \begin{cases} Pr(t_{j'} > p_{i,j} \text{ for } j' = l_i^1, \ldots, l_i^{k-1} \text{ and } t_j \leq p_{i,j} \text{ for } j = l_i^k | L_i, p_i) & \text{if } j \in L_i \\ 0 & \text{if } \; j \notin L_i \end{cases} \tag{1.3}$$

Equation 2.3 shows that students' ranking behavior can be characterized by cutoffs of the matchings. Therefore, expected cutoffs can be used in the examinations of students' behavior.

1.2.3 **Stability**

Stability/Justified-envy is the most commonly used equilibrium notion in non-transferable utility matching markets and also used in preference estimation methods to characterize students' behavior. Since we compare our method with the methods based on stability, we define it in our setting.

Definition 1.2 *Given a matching μ, (i, j) blocks the matching if student i prefers college j to her match and either (1) college j does not fill its quota or (2) college j is matched to another student who has strictly lower priority than student's i. A matching μ is* ***stable*** *if there is no student-college blocking pair.*

Stability can also be defined with college cutoffs. Given a matching μ, a college j is feasible for student i if $t_j(\mu) \leq p_{i,j}$. Feasible colleges for student i is denoted by $\varphi(p_i, T(\mu))$, where $T(\mu)$

represents vector of cutoffs[16] in matching μ. Then,

Lemma 1.1 *μ is stable if and only if $\mu(i) = arg\ max_{j \in \varphi(p_i, T(\mu))} u_{i,j}$ for all $i \in S$.*

This lemma is a basic feature of the college admissions model with finite number of students investigated in many papers (Balinski and Sönmez (1999), Sönmez and Ünver (2011)). Azevedo and Leshno (2016) depart from a finite number of students on one side of the matching and analyze the college admissions game with a continuum of students. We adopt the continuum number of students approach to analyze the college admissions model, which is a more suitable representation of national level college admissions.

1.2.4 Convergence of discrete economies to continuum economies

This section presents the convergence of economies with finite number of students to economies with infinite number of students. With this result, we can analyze students' college ranking behavior as individual choice problems.

Following Azevedo and Leshno (2016) and Fack et al. (2017), we consider sequences of finite size economies denoted by $\{F^S\}_{S \in N}$ such that

$$F^S = \{(u_{i,j}, p_{i,j})_{\{i \in S^s, j \in C\}}, \{q_j^S\}_{j \in C}\};$$

1. There are S students in F^S, whose types are i.i.d. draws from $G \times H$.
2. Each college's capacity relative to S remains constant, i.e. $q_j^S / S = \bar{q}_j$ for all j, where $\bar{q}_j$ is a positive constant.

The continuum economy, E, is defined as follows:

1. A mass of students, S, have type space $[0,1]^J \times [0,1]$ associated with probability measure of $G \times H$.

[16] Since there is one to one relation between scores and priorities in a no ties environment, we are using these concept interchangeably. We continue to use score after we define large matching market structure.

2. College j has a positive capacity $\bar{q}_j$ for all $j \in C$.

Azevedo and Leshno (2016) show that discrete economies converge to a continuum economy when there is a unique stable matching. Fack et al. (2017) modify this finding according to the DA setting and show that for cases where students have ordinal preferences. The convergence results can be seen from the following proposition.

Proposition 1.1 ***(Azevedo and Leshno (2016), Fack et al. (2017))*** *Fix $\sigma \in \{\sigma^S\}_{S \in N}$, where σ is a Bayesian Nash equilibrium of F^S, and apply it to the sequence of finite economies $\{F^S\}_{S \in N}$. Then, we have:*

1. *The cutoffs of finite matching game converges to cutoffs of the continuum game,*

2. *The fraction of blocking pairs in the finite economy converges to zero,*

3. *The asymptotic distribution of cutoffs are normally distributed.*

The proof of this proposition can be found in Azevedo and Leshno (2016).[17] The convergence of finite economies to continuum economies result provides a good approximation to large college admissions. In addition, the resulting distribution of cutoff scores allow us to characterize students' college ranking behavior in a simpler setting. Under the continuum economy, students' behavior can be reinterpreted with the expected outcomes of the matching, which are the cutoff scores in our setting. With a known distribution of cutoff scores, we do not have to consider students' behavior in a game theoretical structure. Instead, students' college ranking strategies can be written according to the distribution of college cutoff scores in addition to personal valuations over colleges. Moreover, this is a better representation of students' behavior in large admissions settings because it is expected that past years' information is utilized by students in their strategies. Therefore, we make the following assumption to utilize large admissions approximation in the analysis of students' behavior.

[17] Abdulkadiroğlu et al. (2015b) also consider similar framework and they note that outcomes of the mechanism can be described with cutoffs.

Assumption 1.1 *Students are assumed to generate distributions of cutoff scores for the forthcoming matching realizations using available information from past years' matching outcomes.*

Assuming that generating a distribution of cutoff scores for colleges is a weaker assumption than assuming perfect information about preferences or the distribution of preferences. Since cutoff scores are observables and there are public records from previous matchings, students can approximate the distribution of cutoff scores by using this information. However, preferences are unobservable and making assumption about them is not easily justified. Assumption 1 is justifiable in many centralized college admissions mechanisms[18] where past years' matching outcomes are public information and there is a strong correlation between the cutoff scores throughout years.[19] In these cases, publicly known cutoff scores are public signals for all students and matching outcomes from the past years play benchmark roles in the formation of students' expectation for the forthcoming matching outcomes.

1.2.5 Students' optimal college ranking strategies in the continuum economy

In this section, we characterize students' optimal college ranking strategies in the continuum economy by allowing individual choice heterogeneities. This characterization is a result of the large market setting, where students primarily use public information of past years cutoff scores to generate assignment beliefs.

In the continuum economy, student i's ROL preparation strategy is a function of her personal valuations for colleges, exam score, past years' matching outcomes, and her idiosyncratic characteristics (ξ_i), $\sigma(u_i, \tilde{p}_i, \xi_i, T^{-1}) : [0,1]^J \times [0,1] \to L \in \mathscr{L}$. In this context, student i maximizes her expected utility using her subjective assignment probabilities (assignment beliefs). We allow each student to generate a distribution of cutoff scores according to their understanding of mechanism

[18]It includes the college admissions mechanism in Turkey, where all realized matching outcome are publicly available and distributed to college candidates in the submission guidance booklet.

[19]The correlation of cutoff scores of college admissions' outcomes in Turkey across years are presented in Appendix. The correlation of cutoff scores are found more that 90% in Ghana for postsecondary admissions Ajayi (2015). In addition, the estimated density of demeaned cutoff percentiles presented in figure A.6.

and last year's matching information (T^{-1} i.e. past year's cutoff score).[20] Consequently, students solve their optimization problems with the following strategy;

$$\sigma_i^*(u_i, \tilde{p}_i, \xi_i, T^{-1}) = arg\,\underset{\sigma}{max} \sum_{j=1}^{K_i} u_{ij} \int \Omega_{i,j}(\tilde{p}_i, \sigma; T) dF(T|T^{-1}, \xi_i) - \kappa \tag{1.4}$$

where $\Omega_{i,j}(\tilde{p}_i, \sigma; T)$ is the assignment probability that takes student i's score and cutoff scores of matching.[21] Assignment of students to colleges according to ex-post cutoff scores is explained in equation 2.3.[22] ξ_i, which is student i's idiosyncratic characteristic, changes assignment probabilities through student i's generated distribution of cutoff scores. This factor relaxes symmetric equilibrium play assumption for the students' college ranking strategies.

The presented framework in equation 1.4 provides valuable understanding of students' college ranking behavior. First, students are maximizing their expected utilities according to their valuations over colleges and their subjective assignment probabilities. Second, with some regularity conditions, we can use students' ranking strategies and propose an empirical strategy accordingly. We provide one more assumption to impose a condition on subjective assignment probabilities of students such that the most important input for generating distribution of cutoff scores is the last year's cutoff scores.

Assumption 1.2 *$\int \Omega_{ij}(\tilde{p}_i, \sigma; T) dF(T|t_{j'}^{-1}, \xi_i) \leq \int \Omega_{ij'}(\tilde{p}_i, \sigma'; T) dF(T|t_j^{-1}, \xi_i)$ if $t_j^{-1} > t_{j'}^{-1}$ for all $i \in N$.*

where t_j^{-1} denotes cutoff score of college year in the last year's matching. Assumption 1.2 says that student i's subjective assignment probability for a college that has a higher cutoff score in the last realized matching is lower or equal than a college that has lower cutoff score in the last matching. These 2 assumptions in the large college admissions model allows us to present the following lemma.

[20] Using subjective probabilities in educational choice context is not new after seminal work Manski (2004). Among others, Zafar (2011), and Attanasio et al. (2011) implement subjective expectations in educational choices.

[21] $\tilde{p}_i$ denotes the scores of students rather than their priorities. Since priorities depend on the other students' performance, using entrance exam scores is better way of characterizing students' behavior in the continuum economy

[22] The existence of the maximum is a result of the bounded utilities, but the uniqueness of maximum requires some conditions that make the expected utility vector smooth and concave.

Lemma 1.2 *For each student with a given exam score, if there is a college that a student's subjective assignment probability is zero, then all colleges that have higher cutoff score in the previous year have zero subjective assignment probability.*

Lemma 1.2 provides an environment to narrow down college options according to predicted ranking behavior. According to lemma 1.2, each student has a limit score ($\bar{P}$) that her subjective assignment probabilities are positive until that limit. Then, we two more assumptions to present our main result.

Assumption 1.3 *For all $i \in N$, $0 < \kappa < u_{ij}\ E_i[\Omega_{ij}]$ if $E_i[\Omega_{ij}] > 0$.*

Assumption 1.3 says that the cost of including any colleges are not bigger than expected gain if there is a positive belief.

Assumption 1.4 *$F(T|T^{-1}, \xi_i)$ has continuous and strictly positive density $f(T|T^{-1}, \xi_i)$, on a support $[\underline{T}_i, \overline{T}_i]$.*

Assumption 1.4 is a regularity condition on subjective expectations with the support conditions of the distribution of cutoff scores. According to this condition, some colleges' assignment beliefs are zero for students because their scores are not part of the cutoff domain. In addition, ξ_i is a competent of the distribution for student i and it includes idiosyncratic differences in the generated distributions.

Theorem 1.1 *Under Assumptions 1-4, a student ranks colleges according to her true preference order in her accessible region to maximize her expected utility unless she have restriction to in the list.*

This is the key prediction of our model and we build our empirical strategy on this finding. Even though this seems restrictive in a setting where students are allowed to rank until a limited number, it is a good representation of observed students' behavior in the data that we applied our method. More than 80% of college applications report rank order lists strictly less than allowed number in Turkey. Therefore, we produce estimation strategy according to our data in our empirical framework.[23]

1.3 Empirical Framework

In this section, we first present challenges in students' preference estimation with a data from centralized college admissions that uses Deferred Acceptance Algorithm. Then, we propose our estimation method based on personal consideration sets, which is developed based on predicted students' college rankings in section 1.2.

In a college admissions where students submit restricted lists, the main challenge is estimating students' preferences without treating a preferred but unranked college in ROL as if it is less preferred. Since making comparisons is the main principle for estimating preferences in choice environments, making invalid comparisons are detrimental for reliable estimates. Possible invalid comparisons emerge under truth-telling assumption because students may not include their preferred colleges if they believe that there is no chance for them to matched with these colleges even though these colleges are more preferred than the listed colleges in their ROLs. This is a serious concern for estimation and there is no ideal solution to identify which colleges are not included because of strategic concerns.[24] Moreover, finding the colleges that are not included because of strategic reasons is one of the objective of estimating preferences from reported ROLs.

Our empirical strategy is estimating students' college preferences using collections of colleges that are expected to considered and found to be accessible to be matched by students. In this way, we do not treat colleges that are not included in ROLs because of underlying true preference or strategic concerns as the same way and do not make invalid comparisons. In addition, this strategy allows us to use all listed colleges in ROLs, which is not utilized by the previous estimation methods (Fack et al. (2017), Kirkebøen et al. (2012), and Burgess et al. (2015)).[25] Therefore, we

[23] If students have more colleges that they may want to include their list than limited number, they can only report a subset of colleges that they want to report. These cases create another possibility for making mistaken comparison for estimating preferences and we excluded these cases in our estimation.

[24] Except some very strong preference assumptions (e.g., homogenous preferences) that has no reason to analyze, we cannot make claim for students' true preference orders.

[25] Asymptotic stability based estimation methods develop estimation assuming that matched choice is the favorite

can utilize complete information in students' reported ROLs and eliminate problems related with strategic reported lists.

Implementing our estimation strategy requires to construct the set of colleges that are expected to be considered and found to be accessible by students. We call them consideration sets. To fix ideas for consideration sets, let us consider student i with score p_i. For student i there is a score limit ($\bar{P}_i$) as in the section 1.2 such that all colleges that are lower expected cutoff scores than this limit are considered to be accessible for her. In other words, her subjective assignment probabilities are positive for all colleges that have expected cutoff score less than $\bar{P}_i$ and she believes that she has no chance for all colleges that have higher expected score than $\bar{P}_i$. In our setting, $\bar{P}_i$ depends on student i's score in the exam, her risk attitude, and understanding of the matching mechanism. Therefore, using $\bar{P}_i$ is our strategy to construct a consideration set for student i, which is a collection of colleges that have lower expected cutoff score than $\bar{P}_i$.

Using consideration sets approach enables us to deal with problems result of the truth-telling and equilibrium-play assumptions because it proposes a data-driven method to estimate students' preference without making invalid comparisons. However, there is no direct way to construct consideration sets in empirical setting. Researchers cannot observe students' subjective assignment probabilities and their personal limit scores. Therefore, Implementing consideration set approach requires to generate a procedure to approximate students' limit scores using past years' matching information and students' predicted ranking behavior in section 1.2. In this paper, we are considering the case, where past years' matching outcomes are publicly available and play the most important role in the students' ranking behavior as in our model. By combining past years' cutoff scores information with a revealed choice framework, we can propose a data-driven procedure for constructing consideration sets. As a result, given that students use the last year's cutoff scores as the primary source of information for preparing their ROLs, we can approximate their limit cutoff scores from their reported ROLs. Since it is assumed that students prepare their ROLs based on our model's prediction, students do not include any college which has higher expected cutoff score

choice among the feasible colleges.

than their score limits. In our notation,

$$\forall i \in S \nexists t_j^{-1} \in L_i \; s.t. \; t_j^{-1} > \bar{P}_i$$

where t_j^{-1} is the last year's cutoff score of college j. Recall that L_i denotes ROL of candidate i. The statement above says that there is no student that includes a college j in her ROL if expected cutoff is higher than her upper limit cutoff score $\bar{P}$. Since last year's cutoff scores are the main determinants of students' expectations (Assumption 1.2), all colleges have higher cutoff score than $\bar{P}$ in the last year are expected to be higher than limit score. As a consequence, we can construct consideration sets based on last year's cutoff scores and write that all ranked colleges in ROLs should be smaller than students' limit scores.

$$max\{t_j^{-1}\} \leq \bar{P}_i \quad \forall j \in L_i \tag{1.5}$$

Equation 1.5 shows us to how can we approximate limit scores ($\bar{P}$) using last year's cutoffs and ROLs. Using this inequality, we can construct consideration sets for each student with the colleges that are lower cutoff scores in the last matching than $max\{t_j^{-1}\} \ \forall j \in L_i$. This construction is also compatible with expected utility maximization problem. A student includes college j in her list if it is in her consideration set and the value of assignment is higher than outside option.[26] Under a positive cost (κ) of adding a college in the ROL, the colleges with have same utility with the outside option and zero assignment beliefs are excluded from the ROLs. Since including an additional college decreases expected utility because of the cost, it is not rational to include in ROL.

A comparable approach of creating personal choice sets is also applied by Fack et al. (2017) and Akyol and Krishna (2017). They develop discrete choice estimation framework by using feasible colleges (i.e., colleges have lower cutoff scores than students' scores.) that are created by ex-post cutoff scores. Apart from dependence of a (asymptotic) stability assumption, one disadvantage

[26]In this analysis we only consider students who report ROL with less than allowed number. This is realistic in our data because majority of students do not submit full lists.

of their method is lack of using compete information in ROLs. Since estimation can be done only in feasible regions, colleges that are not in the feasible region cannot be incorporated to estimation, while they are in the reported ROLs. Constructed personalized choice sets, however, contain feasible regions, hence we can say that comparisons in our consideration sets are more general than feasible regions and incorporates complete information from ROLs.

Finally, a drawback of consideration set approach, even students' assignment beliefs are assumed to be known perfectly, is the impossibility of making an analysis with all colleges in the mechanism. Since personal consideration sets exclude some colleges in the estimation process in order not to make invalid comparisons, a complete analysis can not be done by researchers. Lacking of making complete comparisons across colleges is the cost that is paid for reaching reliable results and it is inevitable in these models.

1.4 Model Specification, Identification, and Estimation

We assume that student i's utility from attending college j can be represented by the following random utility model:

$$u_{i,j} = \theta_j - d_{i,j} + X_{i,j}\beta + \varepsilon_{i,j} \tag{1.6}$$

$V_{i,j} = \theta_j + X_{i,j}\beta - d_{i,c}$ is defined as the deterministic part of the utility. $X_{i,j}$ are student college specific attributes, e.g., interactions of students characteristics and college characteristics. $d_{i,j}$ is the distance from student i's city to college j's city. We normalize the effect of distance to be -1, so the magnitude of other coefficients can be easily interpreted in terms of willingness to travel. In addition, for location normalization, $\theta_1 = 0$.

An crucial aspect in our consideration-based estimation, which is similar with stability-based estimation proposed in the literature, is that students make choice/ranking within their personalized choice sets. This may create problems if the choice sets are determined endogenously. However, we assume that student priorities, risk attitudes, and other factor that influence their consideration sets are independent of students' college valuations. In our model this corresponds to $\xi_i \perp \varepsilon_{i,j}$ for all j. Under this condition, personal consideration sets are determined exogenously and identification goes throughly.

After construction of individual consideration sets (CR), students' college preference estimation from their rank order lists becomes a standard ranked order response analysis. In this setting, student i has right to rank K colleges but she does not have to fulfill her list. We do not restrict the number of options in individual consideration sets, however, they have to be bigger or equal than the number of colleges in the ROL i.e. $\forall\, |CR_i| \geq |L_i|$, because of their constructions.

If each student ranks their options truthfully in their corresponding consideration sets and submits L_i, then we can apply ordered choice model to our data. By following the literature (Hastings et al. (2009), Abdulkadiroğlu et al. (2017)), we consider a conditional choice probability structure in logistic distribution[27] form such that;

$$\begin{aligned} Pr(L_i = (l_i^1, \ldots, l_i^K)|X_i;\Theta) &= Pr(u_{i,l_i^1} > \ldots > u_{i,l_i^{K_i}} > u_{i,j'}\ \forall j' \in CR/L_i|X_i;\Theta) \\ &= \Pi_{j\in L_i} = \Big(\frac{exp(V_{i,j})}{\Sigma_{j' \not\succ_{L_i} j} exp(V_{i,j'})}\Big) \end{aligned}$$

where $j' \not\succ_{L_i} j$ indicates that j' is not ranked before j in L_i, which includes j itself and the colleges are not ranked in L_i. This rank-ordered logit model can be seen as series of conditional logit models: one for top-ranked college-major (l_i^1) being the most preferred; another for the second-ranked college-major (l_i^2) being preferred to all college-major except the one ranked first, and so on.

The model is point identified under standard assumptions for rank-order choice models and can be estimated by maximum likelihood estimation with the log-likelihood function:

$$ln\, L(\Theta|\, Z, |L|) = \Sigma_i^S \Sigma_{j\in L_i} V_{i,j} - \Sigma_i^S \Sigma_{j\in L_i} ln\big(\Sigma_{j' \not\succ_{L_i} j} exp(V_{i,j'})\big)$$

[27] The idiosyncratic error term in $\varepsilon_{i,j}$ is assumed to be i.i.d type-I extreme value.

1.5 Application: College Admissions in Turkey

There is a centralized college admissions system in Turkey. Students who wish to pursue higher education take a nationwide college entrance exam (CEE), which is conducted by the Student Selection and Placement Center (CSSP). The sum of students' college entrance exam scores and scaled and weighted high school GPAs determines students' priorities in admissions. After the announcement of the exam score and scaled GPA, a student, who wants to continue student-college matching stage, submits a ROL of up to 24 college-majors to the CSSP.[28] CSSP matches students and colleges using the student proposing Gale-Shapley Deferred Acceptance Algorithm, developed by Gale and Shapley (1962).[29]

College and major choices are simultaneous in the admissions. Majors are divided into 4 main categories according to high school education tracks, which are quantitative, verbal, composite, and foreign languages. Each major belongs to a track and students are prioritized according to their exam scores in the corresponding track.[30] The college entrance exam is held once a year and candidates are asked multiple-choice questions from 5 main topics, which are Mathematics, Natural Sciences, Social Sciences, Turkish Language, and Foreign Languages. These parts generate 4 different scores and form the main part of the admissions criterion. High school GPA is the other part of the college admissions criterion. Students' high school GPAs enter into total score after it is normalized according high school's overall success in college entrance exam. Students from successful high schools have higher high school GPA scores in the total score calculation. Another important admissions rules is the effect of GPAs on total score. It depends on students' education track choice in high school and college. High school GPA score is multiplied by 0.8 if students choose the same educational track for their college majors, while it is multiplied by 0.3 for

[28] Students who do not submit a list either wait for the next years or utilize outside options that can be universities in abroad or labor market.

[29] In our application, we treat the mechanism and data as if students are proposing DA, which is the how procedure works in Turkey. However, Balinski and Sönmez (1999) show that the outcome is college proposing DA in Turkish college admissions under unrestricted lists. Even though the mechanism which we are using for estimation is college proposing DA, under large matching market assumption, it is very hard for students to find beneficial deviations from true partial order of preferences.

[30] Detailed information for Turkish high school institutional structure can be found in Akyol and Krishna (2017).

majors out of their track. Selecting majors from different education tracks lowers students scores compared to students staying in the same track.[31]

1.5.1 Data

For the empirical analysis, we use three data sets. The first data set is obtained from CSSP. It provides students college entrance exam scores, GPAs, and their ROLs if they have submitted. The second data set is a survey study which is conducted in 2005.[32] The survey sample is randomly selected from senior high school students who would take the CEE in 2005. This survey includes questions about socioeconomic, politic, demographic and educational background. The last data set is a list of cutoff scores for college-majors across years and publicly available on the CSSP website.

There are 12829 senior high school students participants in the survey study, but almost half of them do not submit rank ordered lists after the exam. Students do not submit a list either they find their score lower than their expectations and prefer to wait a year to try again or they prefer not to pursue college education after they learn their scores. Since we are interested in students' college preferences, we exclude students who do not submit ROL.[33]

Table 1.1 reports students' characteristics, choices, and outcomes' statistics. 45 percent of the sample do not submit a ROL after they learn their priorities. The average number of choices in ROL is close to 7. After the matching, 44 percent of the sample is assigned to a college and 12 percent of the sample is assigned to their first choices. Table 1.1 also reports summary statistics students' parent education levels, size of living areas, and income levels.

Table 1.2 presents summary statistics of college-majors. Almost half of the college-majors offer two-year degree to their students. Even though admissions of these college majors are through the same mechanism, there are many specific rules that apply graduates from different high schools.

[31]The total score calculated as follows: $Totalscore_i = CEE_i + 0.8 \times shsGPA_i$ if the choice in the same track, $Totalscore_i = CEE_i + 0.3 \times shsGPA_i$ otherwise. The fraction of out of track choices in the reported ROLs is presented in the figure A.1

[32]For detailed information and descriptive analysis see Alkan et al. (2008).

[33]Since priorities are independent of utilities, these exclusions do not create problem.

Table 1.1: College applicants sample in Turkey: Summary Statistics

Variable	Mean	SD	Observation	Min	Max
	Section 1. Basic Characteristics				
Female	0.46	0.50	12828	0	1
Birth Year	1987.19	0.87	12413	1976	1991
	Section 2. College Entrance Exam Scores				
Quantitative	165.85	63.36	12829	0	298.57
Verbal	184.85	64.42	12829	0	294.45
Equal-Weight	179.15	63.59	12829	0	297.42
Foreign Language	13.41	57.72	12829	0	297.13
High school GPA	80.69	13.08	9874	30.67	100.00
	Section 3. High School Fields				
Natural Science	0.35	0.48	12677	0	1
Social Science	0.12	0.32	12677	0	1
Turkish Math	0.30	0.46	12677	0	1
Foreign Language	0.05	0.22	12677	0	1
Others	0.18	0.38	12677	0	1
	Section 4. Choices and Outcomes				
Submit a ROL	0.55	0.49	12828	0	1
Number of choices	6.99	8.30	12828	0	24
Assigned to a college	0.44	0.49	12828	0	1
Assigned to first choice	0.12	0.32	12828	0	1
	Section 5. Parents Education Levels				
M. higher educ	0.11	0.31	12578	0	1
M. high school	0.19	0.39	12578	0	1
M. middle school	0.12	0.32	12578	0	1
M. elementary	0.47	0.50	12578	0	1
M. illiterate	0.11	0.32	12578	0	1
F. higher educ	0.25	0.43	12518	0	1
F. high school	0.25	0.43	12518	0	1
F. middle school	0.15	0.36	12518	0	1
F. elementary	0.32	0.47	12518	0	1
F. illiterate	0.03	0.16	12518	0	1
	Section 6. Living Area				
Village	0.14	0.35	12412	0	1
Town	0.26	0.44	12412	0	1
City	0.60	0.49	12412	0	1
Income	6.89	2.64	11946	1	12

Notes: This table provides summary statistics for survey study conducted in 2005 by Alkan et al. (2008). College exam scores and high school GPA's are matched with unique identifiers based on 2005 data from the Student Selection and Placement Center.

Table 1.2: Colleges in Turkey: Summary Statistics

Variables	Mean	Sd	Obs	Min	Max
	Section 1. College and major attributes				
New Department	0.03	0.18	7564	0	1
Two Year	0.47	0.50	7564	0	1
Night Education	0.25	0.44	7564	0	1
Distance Education	0.00	0.06	7564	0	1
Private	0.17	0.38	7564	0	1
Full Scholarship	0.01	0.10	7564	0	1
Partial Scholarship	0.05	0.22	7564	0	1
	Section 2. Minimum entrance scores				
Quantitative	210.42	107.34	3672	50.00	381.18
Verbal	266.38	78.04	967	50.00	369.97
Equal-Weight	214.70	92.37	2412	50.00	370.84
Foreign Language	324.74	42.45	262	178.97	393.53
	Section 3. Maximum entrance scores				
Quantitative	226.81	109.56	3672	50.00	399.03
Verbal	284.46	74.50	967	50.00	389.30
Equal-Weight	235.09	93.30	2412	50.00	386.15
Foreign Language	347.69	28.36	262	219.47	400.57

Notes: This table provides summary statistics on attributes of college and majors in the college admissions system of Turkey. These data obtained from Student Selection and Placement Center's web site from the 2005 college entrance system guidance booklet.

We exclude these college-majors from our estimation. Private colleges offer 17 percent of all college-majors. Among private colleges, there are majors provide tuition scholarships which are awarded through college admissions mechanism. 1 and 5 percent of college majors provide full or partial tuition scholarship respectively. Night education is a program that offer education in night times. This program aim to increase capacity of colleges by utilizing off time of campuses. Quarter of all college majors are in night education program.

Table 1.2 also provides average cutoff scores of college-majors for the 2004 matching outcome by education track. College-majors are in quantitative track are engineering, medical education and natural sciences. Equal-weight track includes social sciences, and law school is the second in terms of number of college majors. Verbal track includes theology and literature and foreign language track includes foreign literatures and translation studies.

1.5.2 **Estimation results**

In Table 1.3, we present estimation results based on three identifying assumptions. The first column presents estimates students' college preference estimates where students are assumed to rank truthfully. Under truthful-revelation assumption, we treat the listed colleges as the students' most preferred choices. Second column presents estimates under the stability assumption. The third column presents estimates that are obtained with our proposed method; consideration sets approach.[34]

In order to make consistent comparisons between estimates, we reduce our sample to students who are assigned after matching, only report 4 year degree majors, and do not use all spots in the ROLs. These restrictions reduce our sample to 2863. We exclude students who were assigned to any college because stability-based approach cannot include them into estimation. We prefer to use students who only report 4 year degree majors because two-year program admissions have

[34] In Fack et al. (2017), similar preference estimation with endogenous school cutoff scores are also considered in the secondary school choice context in Paris, but there is no significant difference is founded. By comparing the size of the matching games, there are 1.8 millions of students are in college admissions in Turkey, it is expected to reach similar conclusion.

Table 1.3: Estimation Results

	(Θ_T)	(Θ_S)	(Θ_P)
	$(Truth-Telling)$	$(Stability)$	$(Cons-Sets)$
	Panel A. College and students' characteristics		
Same City	0.7	0.85	0.77
	(0.05)	(0.06)	(0.05)
Partial Sch.	-0.64	-0.17	-0.07
	(0.08)	(0.10)	(0.08)
Full Sch.	-0.83	0.94	0.52
	(0.08)	(0.15)	(0.10)
Distance Education	0.88	-0.93	-0.56
	(0.11)	(0.17)	(0.13)
Night Education	-0.82	-0.97	-0.87
	(0.26)	(0.38)	(0.26)
Private College	-2.98	-1.70	0.45
	(0.79)	(1.17)	(1.12)
	Panel B. Selected college fixed effects		
Bogazici	0.34	1.70	1.56
	(0.09)	(0.12)	(0.10)
Galatasaray	-0.78	81	0.59
	(0.14)	(0.23)	(0.17)
Hacettepe	0.32	0.25	0.33
	(0.07)	(0.10)	(0.08)
Istanbul	0.55	0.64	0.61
	(0.08)	(0.10)	(0.08)
ITU	0.62	1.14	0.96
	(0.09)	(0.11)	(0.09)
ODTU	-0.13	-0.21	-0.11
	(0.08)	(0.12)	(0.10)
Bilkent	-0.65.	2.28	0.89
	(0.52)	(0.97)	(0.89)
Koc	-0.06	3.37	1.56
	(0.53)	(0.97)	(0.89)
Bilgi	-0.06	3.26	1.33
	(0.53)	(0.98)	(0.90)
Sabanci	-0.39	2.90	1.04
	(0.53)	(0.98)	(0.89)
Observation	2863	2863	2863

Notes: This table reports the estimates for the college admissions in Turkey, with coefficients on distance being normalized to -1. All estimates are based on maximum likelihood. Estimation results for the remaining college and students' characteristics are presented in table A.12 in Appendix. Estimation results for all college fixed effects are also presented in Appendix. Standard errors are presented in parentheses.

additional criteria on top of the general admissions and cannot be generalized to population. Finally, because behavioral assumption of our econometric framework is for non-complete ROLs submission, we restrict sample with these students.[35]

The results show clear differences between three different estimations which are based on different assumptions. The difference between truthful-ranking with the other two estimates specifically observed from the college fixed effect estimates. Many elite colleges that are highly selected from students with high priority have smaller or negative fixed effects with the truth-telling assumption. For example, the coefficients for Bogazici University is smaller than many other universities' fixed effects under truth-telling assumption. However, students with highest scores in college entrance exam are listed it frequently.[36] Many students who does not have high priorities do not include it in their ROLs, which leads to a low estimates of its fixed effect under truth-telling assumption.

Another interesting result is the effects of universities capacities on the universities fixed effects' estimates under truth-telling assumption. Universities that have larger capacities have larger fixed effects under truth-telling assumption, however these results do not preserved on the other specifications. This result is another significant finding for assuming truth-telling is creating misleading conclusions. Since universities with larger capacities are selected more in ROLs compare to smaller elite universities, their fixed effects estimates are bigger because of truth-telling assumption.

The estimation based on stability assumption is a modified version of the proposed method by Fack et al. (2017) according to our data. We construct feasible colleges according to cutoff scores which are result of the year 2004 matching.[37] We apply ordered data methods to obtain estimates from constructed feasible sets and reordered feasible ranks.[38]

[35]Estimation results with students who complete ROLs is presented in the tables A.15, A.16, and A.17. These results also point out that the estimates are different than each other in a similar way as in the estimation results with students who do not complete their lists.

[36]Bogazici is the one of the most popular colleges among students who has the highest priorities. In Appendix, we prepare a chart for first 100 students' first college choices in each education track.

[37]Since some of students do not submit a ROL with no feasible college, they are not assigned and we can not include them in our estimation.

[38]The proposed method by Fack et al. (2017) do not include ranking information and only consider the selected

As we explained in our empirical framework, we construct students' consideration sets using cutoff scores from 2004 which are public information. These cutoff scores help us to understand each student's range of colleges that they considered and found to be accessible when they are preparing their ROLs. After construction of personal consideration sets, we apply ordered response models within these sets.

The difference between estimates based on stability and personal consideration sets approach can be also easily seen from the estimated coefficients and college fixed effects. Especially, the estimated private college coefficient is different from each other substantially. The differences between college fixed effects estimates are also remarkable. Even though these differences are suggestive for the differences in estimates as the result of identifying assumptions, we need statistical tests to reach conclusive comparisons.

1.5.3 **Testing stability against personal consideration sets approach**

Under different identifying assumptions, we have distinct estimates for the parameters for students' college preferences. This setting provides an opportunity to test stability assumption against to personalized consideration sets by adopting a Hausman-Type specification test.

We know that if every student play equilibrium strategies, the matching outcome is stable. However, if students play non-equilibrium strategies, the matching outcome do not have to be stable. Personalized consideration set approach is less restrictive than stability assumption because it does not require a stable outcome. Consequently, under null hypothesis that students play equilibrium strategies and outcome is stable, both $\hat{\Theta}_P$ and $\hat{\Theta}_S$ are consistent but only $\hat{\Theta}_S$ is asymptotically efficient. Under the alternative, when students do not play equilibrium strategies, the outcome is not stable but they ranked colleges from their personalized consideration sets. Then, only $\hat{\Theta}_P$ is consistent.

In this setting, we can offer a general specification test developed by Hausman (1978) such that

college as in the binary choice framework.

$$(\hat{\Theta}_P - \hat{\Theta}_S)'(\hat{V}_P - \hat{V}_S)^{-1}(\hat{\Theta}_P - \hat{\Theta}_S)$$

where $(\hat{V}_P - \hat{V}_S)^{-1}$ is the inverse of the difference between the asymptotic covariance matrices of $\hat{\Theta}_S$ and $\hat{\Theta}_P$.

The test rejects stability in the favor of personalized consideration sets (p-value < 0.01). The tests based on truth-telling against to both stability and personalized consideration sets are also rejected strongly.

The second way to test stability assumption against to personalized consideration sets is comparing their predictive power. Even though in most of the cases it is not obvious to make prediction about true preference orders of students, there are some specific scholarship cases which are awarded in the matching allows us to generate a prediction power comparisons.[39] In Turkish college admissions, private colleges provide different tuition and scholarship options for their departments and these options are considered to be different slots in the matching. The preference orders are obvious among the college-major slots, everybody prefers slots which have less tuition over slots which have more tuition. This observation helps us to compare predictive power of different estimation methods.

In order to see the fraction of correctly predictions between stability vs personal consideration set approaches, we focus on students who report a college-major department which they need to pay full tuition if they are matched with them, however, they do not report tuition-free versions. For these students tuition-free version of reported college-majors should have been preferred over the ones reported in their lists. Therefore, if we use our estimates to see their predicted valuations, we can compare predictive power of preference estimation methods. According to stability assumption based 41% of the total cases in our sample are correctly predicted such that they are preferred to the highest predictive valuation for the reported list. This rate is 48% for the estimation result based on personal consideration set approach.[40]

[39] A similar approach is first considered in Hassidim et al. (2016). Then, Artemov et al. (2017) develop a robust equilibrium idea.

[40] We also compare the prediction rates in compare to the highest ranked college-major in the reported lists. The

These tests' results show that personal consideration set approach has better performances, even though it requires weaker assumptions. In the following section of this paper, we focus on the results obtained with personalized consideration sets assumption to conduct welfare analysis under different admission criteria.

1.6 Welfare Analysis

In this section, we use our preferred preference estimates to evaluate the effects of different admissions criteria on students' welfare. Admissions criteria determine students' priorities in the matching, consequently they affect not only students' matching outcomes but also students' ranking strategies in the restricted ROL mechanisms. Therefore, comparing different admissions criteria is an important step in the evaluation of centralized college admissions mechanisms. We study the effects of two admissions criteria which are the two separate scores of the college admissions criteria that form the current priority rule:

1. *College exam score*. College exam score is the only determinant of priority indices. A similar admissions criterion is also used in the high school admissions mechanism in Turkey, Akyol and Krishna (2017).

2. *High School GPA*. Priority indices are based on scaled high school GPAs.

The applied admissions criterion in our data is a weighted sum of scores from a national level exam and a scaled high school GPA.[41] We are interested in these two criteria because of differences of their formation periods. High school GPA is the average of many courses that is taken across 3 or 4 years depends on the high school. However, college entrance exam is only one time exam and the scores are subject to include exam day shocks compare with the scaled GPAs. In addition,

rates are closer (50%, 52%) but still personal consideration sets approach's predictive rate is higher. In addition, we compare result prediction power among the students who report complete lists after estimating their preferences. The predictive power of two methods are the same, 36%.

[41]The implementation of this admissions rule can be considered a unique rule except particular cases. For example, a student goes to high school which is specialized for teacher education, takes additional score based on her scaled high school GPA for teacher majors in college.

students may have different strategies to prepare for college entrance exam and high school courses. Comparison of these two version of measurements shed lights on heterogeneity among college applicants' matching outcomes as well as resulting welfare. Therefore, a policy maker may use any possible differences to design admissions rules to create an equitable college admissions.

Since we are working with a sample, simulating matching outcomes are not possible with the estimated college rankings and counterfactual admissions rules. Instead, we use cutoff scores of colleges to determine assignments with different admissions criteria. We have presented that in large matching mechanisms, outcomes can be represented with cutoff scores due to Azevedo and Leshno (2016) in section 1.2. Therefore, welfare analysis with different admissions criteria by using same cutoff scores is reasonable way of creating counterfactual outcomes without complete matching data in large national admissions. However, this can be thought as a partial equilibrium analysis since the change in reporting ROL strategies and their effects on matching outcomes cannot be evaluated under this method.

Table 1.4 shows welfare changes between their original matches with their counterfactual matches with only scaled high school GPAs is used for admissions. Since applied admissions policy is a weighted sum of two admissions criteria that we are interested, comparing results with original criteria and scaled high school GPAs also provide information about matching with only exam scores.

We present welfare changes in 4 categories i.e., gender, age, income categories, and high school types. Even though results are changing in between groups and students, there is a common pattern in the welfare of students if they are matched according to their scaled high school GPAs. This pattern is a result of the difference of score distributions between national level exam and high school GPA. Scaled high school scores are higher on average, which leads to higher welfare gains if we consider that everything is assumed to be constant.

According to the Table 1.4, Panel 1, female students gain less if the admissions criteria only depends scaled high school compare with males. In panel 2, we observe the opposite case for the older college applicants. Their welfare gains from only high school GPA based admissions criteria

Table 1.4: Welfare Analysis

Variables	Mean	Std Deviation	Max	Min
		Panel 1. Gender		
Male	-0.22	1.29	5.40	-7.39
Female	-0.18	1.17	5.61	-7.07
		Panel 2. Age		
1987/08 >	-0.18	1.22	5.61	-7.39
1987/08 <=	-0.24	1.22	5.03	-6.43
		Panel 3. Income categories		
1	-0.50	1.58	3.52	-5.64
2	-0.21	1.32	1.67	-4.12
3	-0.43	1.85	4.90	-4.91
4	-0.10	1.42	3.81	-3.08
5	0.05	1.51	4.31	-4.83
6	-0.04	1.33	5.24	-6.03
7	-0.27	1.23	4.33	-7.39
8	-0.16	1.25	5.40	-5.02
9	-0.35	1.30	5.61	-7.38
10	-0.12	1.23	4.80	-7.07
11	-0.28	1.14	3.17	-6.29
12	-0.15	0.78	3.86	-3.65
		Panel 4. High school types		
Anatolian (S)	-0.23	1.22	5.40	-5.02
Anatolian Vocational (S)	-0.08	0.91	1.67	-1.97
Teacher Vocational (S)	0.07	1.27	5.03	-7.39
Regional High School	-0.11	0.25	0.00	-0.55
Science (S)	-0.23	1.20	2.78	-6.43
Islamic	-0.25	1.53	5.61	-6.11
Vocational	-1.19	1.59	0.00	-3.50
Private Science (S)	-0.24	1.24	4.80	-7.38
Private	-0.27	1.05	2.78	-4.44
Regional	-0.36	0.74	1.48	-2.82
Regional (S)	-0.28	1.35	5.24	-7.07
Total	-0.20	1.23	5.62	-7.39

Notes: This table presents their indirect utility differences of original matched colleges with their counterfactual matches in admissions rule that only uses their scaled high school GPAs. Indirect utility is presented in terms of willingness to travel.

is lower than older applicants. The welfare effects are also changing among different income levels. According to Panel 3, the main welfare gains are low income groups. However, the welfare changes are not monotonically decreasing with income levels.

Panel 4 presents the welfare effects of changing admissions rules across to high school types. High schools in Turkey differentiated in terms of curriculum and student selection mechanisms.[42] At the end of the middle school, students who are willing to continue their education in selective high school have to take national level exam and students and school are matched according the exam score and students' high school rankings. We indicate selective high school types with (S) in the table 1.4. Results in Panel 4 show that there are differences in terms of welfare effects among high schools. Students from regional and private high schools are benefited the most switch from current criterion to scaled high school GPA criterion. For students from selective vocational high schools have the effects are small even opposite for teacher vocational high school. These results suggest that admissions criteria and high school grading policies as well as curriculum should be taken into account for an equitable college admissions.

To sum up, the results suggest that using different performance measure for admissions criterion has differentiated effects on various income categories, students from different high schools, male vs female, and age groups. Therefore, policymakers may use these effects as an additional tool in their hands in order to affect distribution of matching outcomes across different groups apart from affirmative action policies.

1.7 Conclusion

In this paper, we present a novel, practical, and data-driven method to estimate students' college preferences from rank order lists in college admissions data under the popular Deferred Acceptance mechanism. This mechanism is applied in many secondary school and college admissions mechanisms around the world and our method can be readily applied in any of these cases. The proposed method incorporates strategic behavior in students' college rankings with the help of the large ad-

[42] For more detailed information admissions mechanisms of high schools in Turkey, Akyol and Krishna (2017)

missions model and uses public information from past years' matching outcomes to implement the estimation strategy. Since it utilizes available public data, it is more robust than theory-based estimation approaches which depend on equilibrium conditions. Moreover, this method provides a framework to utilize complete information form rank order lists, which is not possible for the proposed methods in the literature.

We apply our method to college admissions data from Turkey. In order to compare our estimation strategy with the proposed methods in the literature, we test our results in different identifying assumptions; truth-telling and stability. Our tests strongly reject truth-telling and stability hypotheses. In addition, personal consideration approach has better performance in the prediction of students' college-major preference orders.

Finally, we use our preferred estimates in the welfare analysis to understand the effects of different admission criteria. The results show that winners and losers differentiate according to admission criteria and students' characteristics. Consequently, policy makers should consider the effects of performance measures in the determination of admission criterion for school and college admissions.

A further structural analysis of the college admissions data to estimate students' college preferences from rank order lists taking endogenous score formation into account would be a our direction of future research. This extension would allow us to benefit from students' college admissions preparation strategies to improve preference estimation methods rather than depending on only reported rank order lists. This is also important in policy design perspective such that estimation results provide insights about the relationship between college admissions preparation and college choice strategies.

CHAPTER 2

An Empirical Analysis of College Admissions with Endogenous Entrance Exam Scores

2.1 Introduction

College admissions are centralized in many countries and a national level entrance exam is a common practice for sorting students.[1] In these mechanisms, the competitiveness of colleges are typically known from the past years outcomes and students have an idea of the minimum threshold scores (cutoffs) they would need to get to be matched with a particular college. Since getting student-college matching depends on exam scores, students with heterogeneous preferences over colleges may pick different strategies, i.e., levels of time and effort to put into the preparation for the exam, according to the competitiveness of their desired/targeted colleges. Therefore, characterizing how students' college preferences affect exam preparation strategies becomes critical for understanding student behavior during the college admissions process. Moreover, the extent to which students' college preferences affect exam preparation strategies has immediate implications for designing better admissions policies (e.g. Affirmative Action and Financial Aid) and the improvements of students' college preferences estimation methods from reported rank order lists.[2]

The existing college admissions literature does not provide a framework to analyze how students' college preferences affect their exam preparation strategies. Currently, researchers either assume that all students have the same preferences and then analyze student score competition to be assigned to better colleges (Hafalir et al. (2016), Bodoh-Creed and Hickman (2016)) or focus on students' rank order lists (ROL) over colleges and analyze the admissions outcomes disregarding how they prepare for the entrance exams (Balinski and Sönmez (1999), Azevedo and Leshno (2016)). Assuming, however, that students are heterogeneous in their personal valuations over col-

[1] China, Greece, Taiwan, and Turkey are the examples of the countries that use an exam-based centralized college admissions. There are also centralized secondary school admissions examples including Amsterdam, Boston, Paris, and New York.

[2] In the centralized admissions students prepare rank order lists to submit to the national center for the student-college matching. These lists show not only their preferred colleges to be admitted but also the preference order among the chosen colleges.

leges[3], then their optimal preparation strategies will depend on both their heterogeneous abilities and the competitiveness of their targeted colleges.[4] As a consequence, students have score formation (exam preparation) strategies in addition to the strategies for submitting ROL of colleges for the student-college matching stage in a college admissions process.

In this paper, I first analyze the effects of students' college preferences on the entrance exam preparation strategies by providing a new model that encompasses strategies over both score formation and the submission of ROLs of colleges. This model predicts that each student has an optimal exam score according to her personal valuations over colleges, ability, and the competitiveness of targeted colleges. Second, I provide an identification framework based on partitioning colleges into sets to understand the effects of students' college preferences on their score formation strategies. In this framework, if there is a competitiveness difference among colleges according to their observable characteristics (e.g. private colleges are less competitive than public colleges), the effect of students' college preferences on exam scores is identified. Since students are optimized according to their personal valuations, abilities and the competitiveness of targeted colleges, their score formation strategies depend on college characteristics, which create differences in the personal valuations, and the competitiveness of colleges when their abilities are controlled for. Following this rationale, I propose a simple empirical procedure with a single college characteristic to test the implications of the partitioning framework. Then, I apply this procedure to data from Turkey's college admissions. The results show that student exam scores are lower if they prefer colleges with have lower cutoff scores.

The large matching market structure allows me to combine the score formation and submission of ROLs of colleges strategies together. I adopt a supply and demand framework for the student-college matching based on the submitted ROLs of colleges and exam scores, which is developed by Azevedo and Leshno (2016). In this framework, student-college matching outcomes can be

[3]A representative case can be seen from students' college rankings in Turkey. Figures A.4 presents students' first ranked choices of first, median and last 100 students in the college admissions.

[4]College entrance exams that I am considering are not designed for IQ measurements. In addition, the effects of different motivations and incentives on test scores are documented in various context by Fryer Jr (2011), Bettinger (2012), Metcalfe et al. (2011), among others.

approximated with a distribution of cutoff scores. Therefore, under a known distribution of cutoff scores, students optimize and submit their ROLs according to this information. Moreover, with the distribution of cutoff scores information, students further optimize their exam preparations by predicting expected utility that corresponds to each score and the cost of obtaining those scores. As a consequence, a college exam preparation period turns to an individual investment decision problem with risky returns.

Examining the effects of students' college preferences on the score formation strategies requires the use of preference information from the ROLs and public information on college competitiveness levels. The partitioned colleges (which are named as college contest sets) according to an observable college characteristic allow me to calculate the fraction of selected colleges in the students' ROLs from a particular college contest set. By this way, analyzing the corresponding observable college characteristic's effect on the score formation strategies is feasible. Since college characteristics are influential factors in the students' valuation differences for colleges, I can test the effect of one dimension of students' college preferences on the score formation strategies. Estimating of these effects on the score formation strategies requires exogenous test results to control abilities. After student characteristics and exogenous test scores are controlled for, the sign and magnitude of the correlation of college choice and the score formation equations' residuals show the relationship between score formation and ROL preparation strategies because the unobservable preference effects are included in the residuals.

I apply the partition-based estimation produce to estimate the effects of students' college preferences on entrance exam scores to data from Turkey's college admissions. In Turkey, there is a centralized and exam-based college admissions.[5] In this mechanism, students first take the college entrance exam and submit their ROLs after they learn their scores. Even though submission of ROLs is after the announcement, the relationship between exam preparation and ROL submission strategies can be estimated. In the application of the contest sets framework, I use basic college characteristic variations like college types, locations and college-majors as the observable differ-

[5]The details of the college admissions in Turkey are described in section 5.

ences between the colleges in order to partition them into sets. The competitiveness variations of these colleges allow me to analyze the preference effects on the score formation strategies. In particular, private colleges have lower cutoff scores than public colleges, majors are differentiated in terms of their cutoffs, and the colleges in big cities have higher cutoff scores than the colleges in smaller cities.

Estimation results for the private college choices show that there is a significantly negative correlation between the fraction of private colleges in the ROLs and the score formation equations' residuals after the high school entrance exam scores, which are exogenous to the college admissions process are controlled for. This result suggests that students who are inclined to choose private colleges put lower effort into the exam preparation since the admissions requirements for the private colleges are lower than the public colleges. Moreover, these effects remain robust even after family income, parent education and accumulated wealth are controlled for. Following the same practice, I find that students who choose competitive college-majors such as medical school or mathematics have significantly positive preference effect to obtain higher scores. On the other hand, this relationship cannot be generalized to all competitive majors. Finally, I estimate that at least 16% of the variation in the college exam scores can be explained by the students' college preferences.

The significant relationship between the score formation strategies and students' college preferences are important for the evaluation of the effectiveness of admissions policies. In particular, affirmative action policies for increasing racial or socioeconomic diversity in the selective colleges are traditionally based on admissions criteria. I discuss preference driven effects on the score formation strategies in the score-based admissions policies. I argue that traditional methods are relatively less effective if there is considerable preference difference between socioeconomic groups and minorities. These discussions suggest that admission policies that are aiming to change college admissions outcomes can be implemented through new policies that change students' college preferences such as financial aid policies.

The rest of the paper is organized as follows: The remainder of the introduction discusses related literature. Section 2 presents the college admissions model in a discrete setting and its convergence to the limit case. In section 3, I introduce the college contest sets. Section 4 presents the empirical framework for the college admissions with entrance exams based on the college contest sets. Description of the college admissions in Turkey, and the results on estimation with the Turkish data are shown in Section 5. Section 6 concludes.

2.1.1 **Related literature**

This paper is related to several active lines of research. First, it is related to college admissions which have been extensively studied in the economic literature. From the seminal paper by Gale and Shapley (1962), two-sided matching theory investigates the properties of college admissions mechanisms. Research motivated by centralized mechanisms such as student placements in college and public school admissions are pioneered by Balinski and Sönmez (1999) and Abdulkadiroglu and Sönmez (2003). Recent works on college admissions based on student performances include Hafalir et al. (2016), Hickman (2013), Bodoh-Creed and Hickman (2016), and Myong (2016). One main difference between my model and these papers is that I allow for heterogeneous valuations over colleges. In this paper, students have personal valuations that affect their college admissions preparation strategies, while in the existing literature personal valuations for colleges are typically assumed to be homogenous in the exam-based college admissions models. Heo (2017) is one notable exception in this literature. She models student behavior with effort competition among students who have differentiated valuations for colleges and abilities in a small decentralized admissions setting. In her model, there are two colleges and effort is a binary decision.

Large matching markets is another related literature because college admissions can be easily described as large matching markets. Azevedo and Leshno (2016) develop a supply and demand framework to analyze stable matchings by assuming a continuum of students. They show that stable matchings have a simple structure and for every stable matching there exists a vector of cutoffs such that each student demands the college she is matched to. The continuum version of the model typically has a unique matching and is a good approximation for real markets. Menzel

(2015b) works on two-sided matching markets with non-transferable utility when the number of market participants grows large and show identification conditions in the large markets. Specifically, Azevedo and Leshno (2016) and Fack et al. (2015) prepare the base large matching market structure of this paper. They do not, however, consider the competition feature among students in the college admissions which is the main focus of this paper.

The large market assumption is also studied in the contest literature. Olszewski and Siegel (2016) con- sider a contest model with heterogenous players and prizes and they show that the equilibrium outcomes of such contests are approximated by the outcomes of mechanisms. Bodoh-Creed and Hickman (2016) analyze affirmative action policies by modeling college admissions as a large contest. Then, Hickman and Bodoh-Creed (2015) identify and estimate structural parameters of the large college admissions contest model. Similarly, large number of students can be considered in aligned with students' college ranking stage but the large matching market with continuum of students assumption for the ROL submission stage provides a sufficient structure for combining the exam preparation with the submission of ROLs.

Empirical research for college admissions and public school choice followed the theoretical studies. In the context of college admissions and school choice, ROLs provide rich data for econometric investigation. Dealing with strategic interactions in the preparation of ROLs is the main identification challenge and requires structural methods. There are recent studies of student preference estimation in the Boston Mechanism (He (2012), Agarwal and Somaini (2016), Hwang (2015) and Casalmiglia et al. (2014)) and in the Deferred Acceptance Mechanism (Fack et al. (2015), Abdulkadiroğlu et al. (2015a), Akyol and Krishna (2017) and Arslan (2018)). The main difference of this paper from these empirical studies is the endogenous score formation strategies. My model and significant results for the relationship between score formation and college ranking strategies show that disregarding the relationship between college choice and preparation strategies causes biased conclusions from estimation results that do not take endogeneity into account.

Motivational differences and incentive effects on student test performance are also examined in the literature under different conditions. Fryer Jr (2011), Bettinger (2012) and Carpena et al.

(2017) find that financial incentives improve test performance and change student behaviors. Metcalfe et al. (2011) and Jalava et al. (2015) show non-financial incentives' effects on student performance and effort. In this paper, I provide additional evidence that motivational differences, which result from heterogeneity in the student preferences for colleges, have impacts on student exam performances.

Other related research focuses on the preference impacts on college admissions policies. Avery and Levin (2010) analyze early admissions in decentralized college admissions and they show that students use early admissions to signal their enthusiasm to colleges. Wu and Zhong (2014), Lien et al. (2016), and Lien et al. (2017) investigate preference submission timing effects on student-college matching outcomes.[6] In this paper, I consider affirmative action (AA) policies which predominantly focuses on admissions procedures of students from different racial or socioeconomic backgrounds. There are recent papers by Hoxby and Turner (2015), Goodman (2016), Hoxby and Avery (2013), and Dillon and Smith (2017) on decentralized college admissions which document that high-achieving and low-income student applications diverge from those of their higher-income counterparts. These findings provide evidence for the heterogeneity in college choice is not only originates from the heterogeneous abilities but also information differences among applicants. Contrary to decentralized admissions, competition information is common knowledge in the centralized admissions mechanisms. Under these conditions, one may expect to see differences in college choices originate from preference heterogeneity which is a result of informational differences of different social groups about the college opportunities. In this setting, I discuss AA policies that aim to affect college and major preferences in addition to traditional policies which are designed with admission criteria. Financial aid programs are good candidates to influence students' college preferences.

Finally, Turkey college admissions data are also studied in other related papers. Krishna et al. (2017) study the central college admissions and analyze the effects of retaking exams. They de-

[6] They find that ex-ante ranking submission increases fairness in terms of ability, college aptitude and also increases match quality between student and colleges. I estimate a preliminary analysis using submission timing policy change in 2000 in Appendix C.

velop and estimate a structural model of exam retaking using data from Turkey's college admissions. Caner and Okten (2010) investigate risk taking behavior in submission of ROLs according to socioeconomic status by looking at students' major preferences at college entry. Caner and Okten (2013) examine how the benefits of publicly financed higher education in Turkey are distributed among students with different socioeconomic status. Saygin (2013) find that gender has significant effects on college choice in Turkey such that female students prefer lower ranked colleges. However, none of these papers investigates the effects of heterogeneity in students' colleges preferences on the exam preparation strategies.

2.2 **Model**

I model the college admissions process in two stages; score formation and submitting ROL stages. In the score formation stage, students obtain college entrance scores, which determine their priorities. Cost of obtaining a score is assumed to be a decreasing function of ability. In the submission of ROL stage students prepare ROLs knowing their priorities, which is commonly investigated in the college admissions literature.

In the following subsections, I summarize the centralized college admissions setup, student ranking behavior when the number of students are finite and show convergence to the limit game. The limit game is defined with the continuum of students and with the same proportions of college capacities as in the finite game. Students' college ranking and score formation strategies as well as the relationship of these strategies are analyzed using the result from the limit game.

2.2.1 **Centralized student-college matching**

College admissions is a classical example of matching markets and matching theory provides a structure to analyze student behaviors. In this section, I consider a student-college matching in which set of students (S) is indexed by $i \in \{1,\ldots,N\}$ and set of colleges (C) is indexed by $j \in \{0,\ldots,J\}$. College 0 denotes being unmatched. Each college j has $q_j > 0$ seats and these are announced before the student scores[7] are announced. C_0 has unlimited seats. I assume that student i derives $u_{i,j}$ utility from assignment into college j. Each personal valuation of college is assumed to have support $[\underline{u}_j, \overline{u}_j] \in [0,1]$ for all $j \in C$. Colleges rank students according to entrance exam scores which transformed into priorities $p_{i,j}$. A college j prefers student i over student j if and only if $p_{i,j} > p_{k,j}$.

After announcement of priorities, each student has an option to submit a rank order list (ROL) of $L_i = (l_i^1, \dots, l_i^K)$. K denotes the length of the ROL and it is assumed to be smaller than the number of colleges in the mechanism i.e $K < J$. l_i^k is student i's k^{th} choice.[8]

The centralized student-college matching is solved by a mechanism that takes students' ROLs and entrance exam scores that determine student priorities. In this paper, I focus on a truncated student-proposing Deferred Acceptance algorithm that is aligned with my data. The algorithm works as follows:

- *Round 1*: Every candidate applies to her first choice. Each college rejects the lowest-ranked students in excess of its capacity and temporarily holds the other students.
- *Round R*: Every student who is rejected in *Round R-1* applies to the next choice on her list. Each college, pooling together new students and those who were held from *Round R-1*, rejects the lowest-ranked students in excess of its capacity.

The process terminates after any Round N when no rejections are issued. Then, each college is matched with the students that it is currently holding. A matching can be defined as in Balinski and Sönmez (1999); it is an allocation of colleges to students such that no student occupies more than one position. Formally;

Definition 2.1 *A matching is a function $\mu : S \to C$ such that $|\mu^{-1}(C_j)| \leq q_j$ for all $j \in C$. If $\mu(i) = C_0$, student i is not assigned any college.*

[7] In college admissions model, college preferences for students are formed by admission criterion and they are called priorities. I normalized priorities using the $[0,1]$ interval.

[8] Fack et al. (2015) analyze the cost of submitting of a ROL according to the number of colleges in ROLs in many cases. In my model, the cost of submitting ROL corresponds to infinity if $|L|$ greater than K.

If a student is rejected in all rounds of the algorithm, she is assigned to C_0 in this model that means she is not assigned to any college.

2.2.2 Student's college ranking behavior

In this student-college matching students have assignment probabilities to be admitted from each college. These probabilities depend on their priorities and submitted ROLs as well as their rivals' priorities and ROLs. If one expresses student assignment probabilities according to priorities and ROLs, their behaviors can be analyzed according to the expected utilities. Following Fack et al. (2015) and Agarwal and Somaini (2016), assignment probabilities are expressed by;

$$\Omega_{i,j}(L_i, p_i; L_{-i}, p_{-i}) \equiv \begin{cases} Pr(\text{i is rejected by } l_i^1, \ldots, l_i^{k-1} \text{ and accepted by } l_i^k = j | L_i, p_i; L_{-i}, p_{-i}) & \text{if } j \in L_i \\ 0 & \text{if } j \notin L_i \end{cases}$$

Given the matching algorithm, $\Omega_{i,j}(L_i, p_i; L_{-i}, p_{-i})$ is either one for some college in C or zero if there is no ties in priorities. Since priorities and preferences are private information, students form their strategies according to joint preference and priority distribution $G(u|p) \times H(p)$, which is common knowledge.

Student i's college ranking strategy is $\sigma_R(u_i, p_i) : [0,1]^J \times [0,1] \rightarrow \Delta(L)$. In the symmetric equilibrium, σ_R^* solves the following problem for every student:

$$\sigma_R^*(u_i, p_i) \in \underset{\sigma}{arg\,max} \ \left\{ \sum_{j=1}^{K} \ u_{i,j} \int \int \Omega_{i,j}(\sigma, p_i; \sigma^*(u_{-i}, p_{-i}), p_{-i}) dG(u_{-i}|p_{-i}) dH(p_{-i}) \right\} \tag{2.1}$$

Existence of pure and mixed strategy Bayesian Nash equilibrium are established by Milgrom and Weber (1985) while uniqueness is not generally true.

Given an equilibrium σ^* from the realization of preferences and priorities, a matching μ_{σ^*} is observed such that the outcome of matching is summarized with the ex-post cutoff of each college:

$$t_j(\mu_{\sigma^*}) = \begin{cases} min\,\{p_{i,j} | i \in \mu(j)\} & \text{if } j \notin \mu(j) \\ 0 & \text{if } j \in \mu(j) \end{cases} \tag{2.2}$$

where $t_j(\mu_{\sigma^*})$ denotes the cutoff of college j. Equation 2.2 shows that if there is no remaining slot after the matching, cutoff is the minimum priority level of the matched students. Zero cutoff means that particular college does not receive sufficient application to meet its capacity. Assignment probabilities can be redefined with cutoffs such that;

$$\Omega_{i,j}(L_i, p_i; L_{-i}, p_{-i}) = \begin{cases} Pr(t_{j'} > p_{i,j} \text{ for } j' = l_i^1, \ldots, l_i^{k-1} \text{ and } t_j \leq p_{i,j} \text{ for } j = l_i^k | L, L_{-i}, p_i) & \text{if } j \in L_i \\ 0 & \text{if } j \notin L_i \end{cases} \tag{2.3}$$

Stability is the standard equilibrium notion in the non-transferable utility matching markets. Stability is a commonly used concept in two-sided matching, is defined in this setting as follows:

Definition 2.2 *Given a matching μ, (i, j) blocks the matching if student i prefers college j to her match and either (1) college j does not fill its quota or (2) college j is matched to another student who has strictly lower priority than student's i. A matching μ is* ***stable*** *if there is no student-college blocking pair.*

Stability can also be defined with college cutoffs. Given a matching μ, a college j is feasible for student *i* if $t_j(\mu) \leq p_{i,j}$. Feasible colleges for student *i* is denoted by $\varphi(p_i, T(\mu))$, where $T(\mu)$ represents vector of cutoffs in matching μ. Then,

Lemma 2.1 *μ is stable if and only if $\mu(i) = arg\,max_{j \in \varphi(p_i, T(\mu))} u_{i,j}$ for all $i \in S$.*

This lemma is a basic feature of the college admissions model with finite number of students investigated in many papers (among others, Balinski and Sönmez (1999) and Sönmez and Ünver (2011)). Azevedo and Leshno (2016) depart from a finite number of students on one side of the matching and analyze the college admissions game with a continuum of students. I adopt the continuum number of students approach to analyze the college admissions model, which is more suitable representation of national level college admissions. In addition, it allows to track students' exam score formation strategies and ranking behavior in a unified framework.

2.2.3 Convergence of discrete economies to continuum economies

This section presents the convergence of economies with finite number of students which is called discrete economies to economies with infinite number of students which is called continuum economies. With this result, I can analyze students' score formation and college ranking strategies in a unified settings because it allows me to characterize these behaviors as individual optimization problems.

Following Azevedo and Leshno (2016), I consider sequences of finite size economies denoted by $\{F^N\}$ such that

$$F^N = \{(u_{i,j}, p_{i,j})_{\{i \in S^N, j \in C\}}, \{q_j^N\}_{j \in C}\};$$

1. There are N students in F^N, whose types are i.i.d. draws from $G \times H$.
2. Each college's capacity relative to N remains constant, i.e. $q_j^N/N = \bar{q}_j$ for all j, where $\bar{q}_j$ is a positive constant.

The continuum economy, E, is defined as follows:

1. A mass of students, S, have type space $[0,1]^J \times [0,1]$ associated with probability measure of $G \times H$.
2. College j has a positive capacity $\bar{q}_j$ for all $j \in C$.

Azevedo and Leshno (2016) show that discrete economies converge to a continuum economy when there is a unique stable matching. Fack et al. (2015) modify this finding according to the

DA setting and show that for ordinal economies[9], which is the economy where the cardinal preferences are replaced with ordinal preferences, converges to the continuum economy. In the ordinal economy, the demand for each college in $E_{\tilde{\sigma}}$ as a function of the cutoff is:

$$D_j(T,\tilde{\sigma}) = \int\int 1(u_j = max_{j' \in \varphi(p_i,T) \cap \tilde{\sigma}_R(u_i,p_i)}\, u_{j'})dG(u_i|p_i)dH(p_i)$$

where $\tilde{\sigma}_R(u_i, p_i)$ denotes the set of schools ranked by student i as in the cardinal economy. Then, convergence results can be seen from the following proposition.

Proposition 2.1 ***(Azevedo and Leshno (2016), Fack et al. (2015))*** *Fix $\tilde{\sigma}_R \in \{\sigma_R^N\}_{N\in\mathscr{N}}$, where $\tilde{\sigma}_R$ is a Bayesian Nash equilibrium of F^N, and apply it to the sequence of finite economies $\{F^N\}_{N\in\mathscr{N}}$. Then, I have:*

1. $sup_{J\in\mathscr{N}}||T(\mu_{(F^N,\tilde{\sigma}_R)}) - T^\infty|| \to 0$,

2. *The fraction of blocking pairs in the finite economy converges to zero,*

3. *If $E_{\tilde{\sigma}_R}$ has a C^1 demand function and $\partial D(T^\infty, \tilde{\sigma}_R)/\partial T^\infty$ is non-singular, the asymptotic distribution of cutoffs in F^S is*

$$\sqrt{N}(T(\mu_{(F^N,\tilde{\sigma}_R)}) - T^\infty) \to Normal(0, V(\tilde{\sigma}_R))$$

where T^∞ is the cutoff vector in E, $V(\tilde{\sigma}_R) = \partial D(T^\infty, \tilde{\sigma}_R)^{-1}\Sigma(\partial D(T^\infty, \tilde{\sigma}_R)^{-1})'$ and

$$\Sigma = \begin{pmatrix} \bar{q}_1(1-\bar{q}_1) & -\bar{q}_1\bar{q}_2 & \dots & -\bar{q}_1\bar{q}_J \\ -\bar{q}_2\bar{q}_1 & \bar{q}_2(1-\bar{q}_2) & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ -\bar{q}_J\bar{q}_1 & \dots & \dots & \bar{q}_J(1-\bar{q}_J) \end{pmatrix}$$

[9] I define ordinal economies as economies where students have ordinal preferences. Cardinal economies are also defined analogously, they are economies where students have cardinal valuations and they are representing their preferences.

The proof of this proposition can be found in Azevedo and Leshno (2016).[10] The convergence of finite economies to continuum economies provides a good approximation to real markets. Under this framework, whenever the continuum economy has a unique stable matching, all stable matchings of the discrete economies converge to a unique stable matching of the continuum economy. Hence, all stable matchings of the large economies becomes similar.[11]

The large matching market with a continuum of students allows me to consider student behavior in the college admissions game in a simple setting. Under the continuum economy, student behaviors can be reinterpreted with the distribution of cutoff scores. With a known distribution of cutoff scores, students do not have to consider rivals' strategies in a college admissions process. Instead, they optimize according to the distribution of college cutoff scores in addition to their abilities and personal valuations over colleges. Therefore, to utilize this result in analyzing students' behavior in a college admissions process I make the following assumption.

Assumption 2.1 *Students are assumed to know the distribution of college cutoff scores,* $F(\mathbf{T})$.

Even though the theoretic result and my assumption simplifies the analysis, it does not have to represent real life applications. On the other hand, Assumption 2.1 is justifiable in many centralized college admissions mechanisms[12] where past matching outcomes are public information and there is a strong correlation between the cutoff scores across years.[13] Publicly known cutoff scores is decisive information because cutoff scores from the past years play benchmark role for the forthcoming matching outcomes. This assumption is crucial for students' preparation period optimizations in this analysis. Since a student-college matching with a large number of students is approximated with a distribution of cutoff scores, I can use this information to analyze students' entrance exam preparation strategies.

[10] Abdulkadiroğlu et al. (2015b) also consider similar framework and they note that outcomes of the mechanism can be described with cutoffs.

[11] I assume that there is a unique stable matching in the empirical part.

[12] It includes the college admission mechanism in Turkey, where all realized matching outcome are publicly available and distributed to college candidates in the submission guidance booklet.

[13] The correlation of cutoff scores of college admission in Turkey across years are presented in Appendix. In addition the estimated density of demeaned cutoff percentiles presented in figure A.6

2.2.4 **Optimal college rankings in the continuum economy**

Under the continuum economy with a known distribution of cutoff scores, students' college ranking behaviors are characterized as individual optimization problems. Note that priorities in the continuum model are replaced with scores ($\tilde{p}$) because priorities depend on rivals' performances. Student i's strategy to prepare her rank order list, which is a function her personal valuations over colleges, exam score and the distribution of cutoff scores, $\sigma_R(u_i, \tilde{p}_i; \mathbf{T}) : [0,1]^J \times [0,1] \to L \in \mathscr{L}$. Consequently, students with their valuations over colleges and scores solve the maximization problem with the following strategy;

$$\sigma_R^*(u_i, \tilde{p}_i; \mathbf{T}) = arg\,\underset{\sigma}{max} \sum_{j=1}^{K} u_{ij} \int \Omega_{i,j}(\tilde{p}_i, \sigma; \mathbf{T})\, dF(\mathbf{T})$$

where $\Omega_{i,j}(\tilde{p}_i, \sigma; \mathbf{T})$ shows the assignment probability of student i's for college j according to her score, rank order list, and cutoff scores. Assignment of students to colleges with cutoff score explained in equation 2.3.[14]

The other advantage of the large market approximation is making characterization of students' preparation period strategies possible. Since the matching outcomes of the continuum economies can be represented with the distribution of cutoff scores and the distribution information is consulted in the students' college ranking decisions, students additionally optimize their entrance exam score formation strategies accordingly.

2.2.5 **Student score formation behavior in the continuum economy**

Students' score formation strategies can be incorporated into the college admissions process with the known distribution of cutoff scores. The score formation strategies therefore represent student preparation behaviors for college entrance exams in the large matching market setting. In this model, students put costly time and effort to maximize their college admissions outcomes that

[14] Existence of the maximum follows from bounded utilities, but the uniqueness of maximum comes with regularity conditions that make the expected utility vector smooth and concave.

was described in section 2.4. In addition to assumptions in the previous section, heterogenous abilities among students is assumed. Utilities and abilities are private information and i.i.d. draws from the joint distribution $\xi(\mathbf{u}, a)$.

With the continuum characterization of the matching game, I now consider the college admissions outcomes as a distribution of cutoff scores. Under this framework, the problem of a student i is to choose her college entrance exam score. The large matching approximation leads that student i's college entrance exam score formation decision is independent of her rivals. In this case, student i's score formation strategy is a function of her valuations over colleges, ability and the distribution of cutoff scores $\sigma_P(u_i, a_i; \mathbf{T}) : [0,1]^J \times [0,1] \to \tilde{p} \in \tilde{\mathscr{P}}$,

$$\sigma_P^*(u_i, a_i; \mathbf{T}) = arg\,\underset{\tilde{p}}{max} \sum_{j=1}^{K} u_{ij} \int \Omega_{i,j}(\tilde{p}_i, \sigma_R^*(\tilde{p}_i, u_i); \mathbf{T})\, dF(\mathbf{T}) - \kappa(\tilde{p}_i; a_i) \tag{2.4}$$

where $\Omega_{i,j}(\tilde{p}_i, \sigma_R^*(\tilde{p}_i, u_i); \mathbf{T})$ is assignment mapping that is explained above. $\kappa(\tilde{p}_i; a_i)$ is the cost function and represents cost of obtaining score $\tilde{p}_i$ according to a student with ability a_i. Since cutoff scores are not known and the matching outcome can be approximated with a normal distribution as it is known from the section 2.3, students optimize themselves according to this information. Therefore, student i chooses her optimal score $\tilde{p}_i^*$ solving equation 2.4 without considering her rivals' strategies.[15]

The equation 2.4 is interpreted as a prospective student's score choice in the college admissions process when she knows the distribution of cutoff scores and the disutility that she bears to obtain scores. As a consequence, agents' strategies in the large college admissions process like a national level centralized admissions can be approximated by individual problems rather than players' strategies depend on their rivals' as in contest or all-pay auction models.

[15]Bodoh-Creed and Hickman (2016) consider also college admission as a large contest and claim that students college admission preparation effort are decided by following previous years college admission information.

2.3 Identification by partitioning colleges: College contest sets

The college admissions model in section 2 shows that personal valuations over colleges are one of the determinants of the optimal scores in the college admissions with entrance exams. In this section, I develop an identification method for the effects of personal valuations over colleges on the score formation strategies by partitioning colleges into mutually exclusive sets, which are named as college contest sets. The college contest set is described as a partition of colleges according to an observable common feature of colleges (e.g. private colleges or public colleges). These sets allow me to characterize competitiveness differences of colleges according to observable college characteristics. Since the distribution of cutoff scores are known by assumption in my model, the competitiveness differences between colleges are understood from students. With this structure, I quantify the selected colleges in students' ROLs from a particular college contest set and analyze the corresponding observable college characteristic's effect on the score formation strategies. Since college characteristics are influential factors in the students' valuation differences over colleges, I can test the effects of one dimension of students' college preferences on their score formation strategies.

To clarify the college contest set framework and its relationship with the score formation strategies, consider a simple example in a centralized college admissions mechanism where there is simultaneous college and major choice. The majors are differentiated in terms of competitiveness and students have different preferences for each major.

Example 2.1 *Assume that there are two students s_1 and s_2 with equal abilities who differ in terms of their major preferences. There are two colleges offers two majors m_1 and m_2. Students are allowed to rank up to two college-majors in their ROLs. The major is the only factor in the student valuation differences for colleges. Student personal valuations for college-majors as follows:*

	Student 1	Student 2
m_1	100	60
m_2	50	90

Each college-major has cutoff score in the [0,100] interval and they have the following distribution of cutoff scores:

	Mean	SD
m_1^1	70	9
m_1^2	71	10
m_2^1	50	10
m_2^2	49	11

Two students have the same cost of obtaining any entrance exam score because they have the same abilities. Student 1 and 2 may maximize their utilities from the expected matching outcomes with different scores because their major preferences make them to compete in different levels. With the given distribution cutoff scores information, there is a higher chance of student 1 maximizes her expected utility with a higher score than student 2's because student 1 prefers major 1 that has higher expected cutoff scores. The cost of obtaining higher score can be compensated with her personal valuation for the major 1. In this example, student 2 prepares the exam according to distribution of major 2's cutoffs and submits (m_2^1, m_2^2) in any case. Student 1, however, needs to choose her preparation strategy and submits ROLs accordingly. If the cost of obtaining higher score is compensated by her valuation differences of the two majors, she prepare more and submit (m_1^1, m_1^2). Therefore, there is a higher chance to observe two different scores with two different ROLs from the students with same abilities if there is substantial differences in valuations over colleges for student 1. Based on this example, if there is a measure to control students abilities that is exogenous to college admissions process, the effects of preferences on entrance exam preparation strategies from exam scores and submitted ROLs.

To formalize the contest choice set framework, consider a collection of contest sets Z and which are indexed by $z \in \{1, \ldots, Z\}$. College contest sets are mutually exclusive collections of colleges. Each contest set has competition level ($\tilde{t}$), which are aggregate information level that comes from the distribution of college cutoff scores. Each college contest set z includes at least one college and a college is allowed to be in only one college contest, i.e. $c_j \in z_k$ then $c_j \notin z'_k$ $, k \neq k'$. This can be expressed that each college j belongs to one z and $\cup_j c_j^z = C$, where c^z denotes colleges that are in the contest set z. For example, there are two college contest sets in terms of college types; private college and public college contest sets.

In this framework, it is assumed that students have complete information for their personal valuations over colleges and understand the differences in the competitiveness between college contest sets which comes from the distribution of cutoff scores differences of the elements of the contest sets. The competition level $\tilde{t}$ is the expected minimum success level such that if a student have a score more than the minimum level, she is assigned to a college in admissions context. In these conditions, students choose contest(s) according to college contest characteristics in the beginning of college admissions process so that they optimize entrance exam preparation strategies according to the benefits of obtaining a score. A contest choice of student i under perfect competition level information is represented as;

$$z_i^* = arg\,\underset{z}{max}\, U^i(\tilde{p}(z), z) - \kappa(\tilde{p}(z), a) \tag{2.5}$$

where $U^i(\tilde{p}(z), z)$ is the utility function of student i that takes score $\tilde{p}$ and contest z as inputs. The utility represents the value of being assigned with a college in contest z for student i. Note that students' score formation strategies depend on the contest choice. A student i may choose two different scores for two different contests $\tilde{p}_{z_i} \neq \tilde{p}_{z_j},\ i, j \in Z,\ i \neq j$ because competition and valuations for colleges differ her optimization strategies.

In the college admissions model, students only know the distribution of college cutoff scores as competition level information. The distribution of cutoff scores also provides the expected competitiveness information for the contest sets. For example, if private college cutoff scores are lower than public college cutoff scores on average, then in the contest level, private college contest is not as aggressive as public college contest. Hence, contest choices of students depend on the expected competitiveness as presented in the equation 2.6.

$$z_i^* = arg\,\underset{z}{max}\, E_{c_z}(U^i(\tilde{p}(z), z) - \kappa(\tilde{p}(z), a)) \tag{2.6}$$

where E_{c_z} denotes the expectation with respect to contest competition levels.[16]

2.3.1 Assumptions

I adopt standard assumptions for the contest choice model. There is one main objective of these assumptions; establishing the existence of individual optimization in the centralized college admissions process.

Assumption 2.2 *For each* $(u,a) \in [\underline{u},\overline{u}] \times [\underline{a},\bar{a}]$, $\kappa_p(\tilde{p},a) > 0$, $\kappa_a(\tilde{p},a) < 0$ *, and* $\kappa_{pp}(\tilde{p},a) > 0$.

Assumption 2.3 *Contest utility function U satisfies vNM utility function properties.*

Assumption 2.4 *For each i and z,* $U^i_p(\tilde{p},z) \geq 0$, $U^i_{pp}(\tilde{p},z) \leq 0$.

Assumption 2.2 states that the cost function is strictly increasing in admission score and it is strictly convex. Moreover, it is decreasing with ability. Assumption 2.3 provides regularization for the expected utility function. Assumption 2.4 states that the contest utility function is weakly increasing and weakly concave. These assumptions imply that students within contest decision problems have global maximums.

Theorem 2.1 *In a centralized college admissions with entrance exam where students have personal valuations over colleges and abilities, under the assumptions 2.1-2.4 each student has an optimal score for each choice of college contest set.*[17]

Given personal valuations over colleges and abilities, students choose contest set(s) and optimize college entrance exam preparation strategies according to this choice(s). Therefore, chosen contest sets are important in the college entrance exam preparation strategies. Theorem 2.1 can be translated into the following corollary in terms of contest set choice and the college entrance exam preparation strategies.

[16]In my framework, it is aggregation of competition level information from colleges in the contest. Al-Najjar and Pomatto (2016) study choice under aggregate uncertainty which represents a similar case of the aggregation of colleges in contests.

[17]The existence of optimal score in Theorem 1 is standard under the given functional assumptions.

Corollary 2.1 *The competitiveness of the chosen contest sets determine college entrance exam preparation strategies; i.e. if a student prefers to be matched with colleges in a contest set which have lower expected cutoff scores than other contest sets, the incentive for obtaining a higher score is lower because of time and effort cost.*

The Corollary 1 allows me to construct a testable implication in the centralized college admissions with entrance exam. Under the large matching market structure, the distribution of cutoff scores provides the competition information. In addition, students are atomistic in terms of their effects on the competition structure of colleges under large markets. Thus, I can generate an empirical framework based on Corollary 2.1 using public level competition information, students' exam scores, submitted ROLs and exogenous test scores.

2.4 Econometric Framework

This section develops an econometric method based on students' college choices from college contest sets. The data that I am considering in this paper consist of rank order lists (ROL), student college entrance scores, and other test scores that are exogenous to college admissions process. I first present construction of contest choice variables which are either the fraction of college choices in the students' ROLs from the college contest sets or an indicator variable according to majority of colleges. Second, I discuss the advantages and disadvantages of contest set framework in the analysis. Then, econometric specification to estimate the relationship between score formation strategies and ROL submission strategies is presented.

2.4.1 Contest choice variable

The main objective in the construction of contest choice variable is quantifying student college choices from different contest sets. ROLs are observable elements for student preferences for colleges to work on. They are sets of ordered choices of each student's decision for student-college matching and submitted to the national admission center.[18] Since college choice and ranking in this setting is not one choice, I offer three type of contest choice indicators and denote them with f, $\tilde{f}$, and $\tilde{b}$. With these variables, I can quantify the fraction or the majority of colleges from different contest sets in one variable. In addition, I can take the order information into account. The construction of the contest choice indicators as follows;

$$f_z^i = \frac{\sum_{j=1}^{K_i} 1\{l_i^j \in z\}}{K_i}$$
$$\tilde{f}_z^i = \frac{\sum_{j=1}^{K_i} w_j j\, 1\{l_i^j \in z\}}{K_i}$$
$$\tilde{b}_z^i = 1\left\{\frac{\sum_{j=1}^{K_i} w_j j\, 1\{l_i^j \in z\}}{\sum_{j=1}^{K_i} w_j j} \geq 0.5\right\}$$

where f_z^i denotes the fraction of student i's ROLs that is filled with colleges in the contest set z. Recall that, l_i^j denotes student i's ROLs j^{th} choice and K_i denotes that number of listed colleges in the ROL. $\tilde{f}_z^i$ is the weighted fraction of student i's ROLs filled with colleges that is in the contest set z according to the order of colleges in the ROL.[19] b_i^z is a binary choice variable which becomes 1 if the weighted fraction of selected colleges from contest set z is bigger or equal than 0.5. With these formulations of contest choices, I can construct various choice fractions and binary choices from the ROLs.[20]

2.4.2 Discussion: Advantages and disadvantages of contest sets method

Before presenting the econometric specification with contest choice variables, there are few issues to discuss with the selection of the econometric methodology. First, the estimation of stu-

[18]One feature of rank order list is the property of lowering the exam shock effects. Even though ROLs are truncated by assignment mechanisms, they can cover some of the colleges those cutoff scores are around the targeted score. Thus, likelihood of observing targeted colleges increase from researchers.

[19]Even though order of ROLs show preference relations in the list, in f's formulation I do not consider these relations. This approach is supported if selected colleges in ROLs are similar enough in terms of preference. In $\tilde{f}$ formulation, I weight the order with Borda count method and generate weighted fraction.

[20]An example for the constructing contest choice variable is presented in the Appendix D.

dent preferences for colleges can be estimated with ordered choice (OC) responses in principle. However, implementation in OC methods with endogenous regressors and high number of choices is computationally very hard problems.[21] The second issue is the interpreting student's preference effect on college entrance exam preparation strategies under OC framework. Because the heterogeneity in the choices and the complexity of the preference effects, it is hard to reach conclusive results. On the other hand, contest sets allows me to separate the effects of students' preferences on preparations strategies in a simple setting. Consequently, a simple estimation and cleaner interpretation for the effects of students' preferences over colleges on preparation strategies are the main advantage of contest sets framework over ordered choice responses.

Contest sets framework does not provide a general identification and estimation methodology for students' college preferences. The preference estimation is an important step for counterfactual analysis and policy design. Lack of a general method is the main disadvantage of contest sets framework. On the other hand, contest sets provide a simpler methodology to test for the existence of the effects of students' preferences on admissions preparation strategies. This provides a crucial evidence to generate estimation methods for student preferences in the college admissions with endogenous scores.

2.4.3 Model specification

As the fraction of colleges[22] from the contest set z in the ROL is the key quantity of interest for indicating contest choice, I investigate the relationship between college entrance exam preparation strategies and contest choice from college admissions data. According to Theorem 1, students choose colleges that are elements of different contest sets to maximize their expected utility of the admissions outcome minus the cost of preparing for the exam. If I separate colleges into two contest sets, which are differentiated in terms of competitiveness, the fractions of colleges

[21] Among others, Jagabathula and Rusmevichientong (2016), Alptekinoğlu and Semple (2016) discuss that computational problems in ordered choice data with endogenous regressors that considers assortment and pricing decisions which is an analogous problem of college admissions with endogenous scores.

[22] Since I can construct binary and fractional responses in the same specification, it is safe to consider one case.

from contests sets show students preference intensities for colleges from two different contest sets. Therefore, the aim of this section is proposing an inference procedure to evaluate how fractions of chosen colleges affect the college entrance exam preparation strategies when student characteristics are controlled for.

Note that timing of the events in the college admissions mechanism is crucial to choose the correct econometric specification. In the college admissions model that is considered in this paper, submission of ROLs occurs after the announcement of examination scores.[23] Learning college entrance exam scores is im- portant for two reasons. First, exogenous exam day shocks can be seen after score announcement and ranking strategies can be updated with the new information. Second, even though ability plays an important role in score formation, after announcement of examination scores, students optimize their ROLs independent of their abilities. In other words, a student prepares herself for the exam by optimizing over her ability, personal valuations over colleges to obtain a score. After score announcement, realized scores, student college preferences and distribution of cutoff scores are the determinants of the submission of ROL strategies.

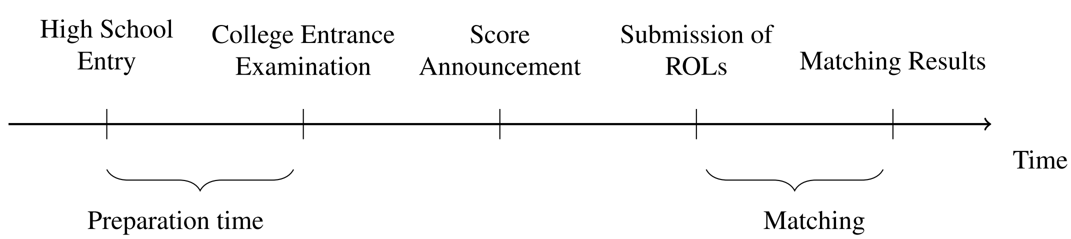


Figure 2.1: College admission process timeline

To model the contest choice indicator f, which is a fractional response variable, I use the following specification, in which the conditional expectation rate is given by

$$E(f|x_1,\tilde{p},\eta) = \Phi(\alpha + x_1\beta_1 + \beta_2\tilde{p} + \eta) \tag{2.7}$$

[23]Bond et al. (2016) study the effects of positive and negative shocks on human capital investment using data from SAT test realizations. They find that positive shocks cause students to choose more selective colleges that charge higher tuition and have higher graduation rates.

where Φ is the cumulative distribution function (CDF) of the standard normal distribution. x_1 captures observable student characteristics and $\tilde{p}$ denotes normalized college entrance exam score. η represents student preference factor. To estimate the parameters $\theta = \{\alpha, \beta_1, \beta_2\}$ in an hypothetical world where the student preference factor is part of the idiosyncratic error, I can use the quasi-maximum likelihood (QML) method proposed by Papke and Wooldridge (1996). Gourieroux et al. (1984) shows that the resulting QML is robust as it yields a consistent estimator whenever the conditional mean is correctly specified, and is efficient for a class of generalized linear models (GLM). However, η in the equation 2.7 is an unobservable element which corresponds to one component of the student preference in my framework. It has been shown in the choices from the contest sets that preferences play a notable role in the determination of college exam score, $\tilde{p}$. Under this condition, I need to deal with the endogeneity in the college exam score formation. To model college entrance score ($\tilde{p}$) formation, I use the following reduced form specification:

$$\tilde{p} = \delta + x\gamma + \varepsilon \tag{2.8}$$

The set of regressors x_1 in the equation 2.7 is strict subset of the regressors x in the equation 2.8 which says that there is at least one exogenous regressor in the score formation equation. I follow Wooldridge (2010) and Wooldridge (2014) for a fractional response model with an endogenous explanatory variable (EEV). There are two models are suggested for consistent estimation; control function approach based on Rivers and Vuong (1988) and joint QML estimation (QMLE) framework. Note that the construction of fractional response variables is designed to indicate the effect of one dimension of the students' college preferences. Students' preference components are assumed to be independent of each others. Therefore, I am assuming the other preference components are the part of the idiosyncratic error term.

In the control function (CF) approach, the first step is to obtain OLS residuals $\hat{\varepsilon}$ from regression college entrance exam score ($\tilde{p}$) on student characteristics (x_1) and high school entrance exam score (x_2). Then, I can use fractional response model based on Papke and Wooldridge (1996) of college choice from a contest set f on $\tilde{p}$, x_1, and estimated residuals $\hat{\varepsilon}$ to estimate scaled coefficients. I denote the coefficient of $\hat{\varepsilon}$ with ω.

To implement joint QMLE specification, I follow the fractional response model by Wooldridge (2014). The fractional response with EEV leads the same log likelihood as the probit model specification of f, even though f is a fractional response.[24] Then, under the distributional assumption, I can estimate college entrance exam score formation and choice equations using joint QMLE methods with log-likelihood function[25]:

$$\max_{\theta_1,\theta_2} \sum_{i=1}^{N} [q_1(f_i, \tilde{p}_i, z_i, \theta_1, \theta_2) + q_2(\tilde{p}_i, z_i, \theta_2)]$$

where $\theta_1 = (\alpha_\eta, \beta'_{1\eta}, \beta_{2\eta}, \rho)'$ and $\theta_2 = (\delta, \gamma', \tau^2)'$ [26]

$$\begin{aligned} q_1(f,p,x,\theta_1,\theta_2) =& f\ \left\{ log\ \Phi[\frac{\alpha_\eta + x_1\beta_{1\eta} + \tilde{p}\beta_{2\eta} + (\rho/\tau)(\tilde{p} - \delta - z\gamma)}{(1-\rho^2)^{1/2}}]\right\} \\ &+(1-f)\ log\ \left\{1 - \Phi[\frac{\alpha_\eta + x_1\beta_{1\eta} + \tilde{p}\beta_{2\eta} + (\rho/\tau)(\tilde{p} - \delta - x\gamma)}{(1-\rho^2)^{1/2}}]\right\} \\ q_2(p,x,\theta_2) =& -log(\tau^2)/2 - (\tilde{p} - \delta - x\gamma)^2/(2\tau^2) \end{aligned}$$

where τ^2 is the variance of college entrance exam score. The main objects of interest are the parameter ω, ρ and the coefficients of college exam score $\tilde{p}$. ω and ρ represent the correlation of the unobservable parts from the college contest choice and college entrance exam score formation equations. Under two econometric specifications, the correlation comes from the one component of unobserved students' preference factor. Therefore, any significant result for ω and ρ captures the relationship between college preference and college entrance exam formation and provides a test for the relationship.

[24] I present the derivation of equality in the appendix.

[25] Obtaining of joint distribution of $(f, \tilde{p}|x)$ is presented in Appendix E.

[26] The parameters in the fractional response variable equations are identified up to scale. $\alpha_\eta, \beta'_{1\eta}, \beta_{2\eta}$ represents scaled versions of $\alpha, \beta'_1, \beta_2$. I refer the reader to Wooldridge (2014) for detailed identification and estimation under these conditions.

2.5 Application: College Admissions in Turkey

There is a centralized college admissions system in Turkey. Students who wish to pursue higher education take a nationwide college entrance exam (CEE), which is conducted by the Student Selection and Placement Center (CSSP). The sum of students' college entrance exam scores and scaled and weighted high school GPAs determines student priorities in the admissions. After the announcement of the exam score and scaled GPA, a student who wants to continue student-college matching stage submits a ROL of up to 24 college-majors to the CSSP.[27] CSSP matches students and colleges using the student proposing Gale-Shapley Deferred Acceptance Algorithm, developed by Gale and Shapley (1962).[28]

College and major choices are simultaneous in the admissions. Majors are divided into 4 main categories according to high school education tracks, which are quantitative, verbal, composite, and foreign languages.[29] Each major belongs to a track and students are prioritized according to their exam scores in the corresponding track.[30] The college entrance exam is held once a year and candidates are asked multiple-choice questions from 5 main topics, which are mathematics, natural sciences, social sciences, Turkish language, and foreign languages. These parts generate 4 different scores, which are the main part of the admission criterion for each track. High school GPA is the other part of the college admissions criterion. Students' high school GPAs enter into total score after it is normalized according high school's success in the college entrance exam. Students from successful high schools have higher high school GPA scores in the total score calculation. Another important admissions procedure in the GPA criterion is based on students' education track choice in the high school. High school GPA score is multiplied by 0.8 if students choose the same educational track for their college majors, while it is multiplied by 0.3 for majors out of

[27] Students who do not submit a list either wait for the next years or utilize outside options that can be universities in abroad or labor market.

[28] Balinski and Sönmez (1999) show that the outcome is college proposing DA even though candidates submit ROLs.

[29] Most college majors are in quantitative track which includes engineering, medical school, and math. The composite track includes social sciences and law is the second most in terms of number of college majors. The verbal track includes theology and literature and the foreign language track includes foreign literatures and translation majors.

[30] Detailed information for Turkish high school institutional structure can be found in Akyol and Krishna (2017).

their track. Selecting majors from different education tracks lowers students scores compared to students staying in the same track[31] if the choice is in the different track.

2.5.1 Data

For the empirical analysis, I use three data sets. The first data set is obtained from CSSP. It provides students college entrance exam, high school entrance exam scores and GPAs, as well as their ROLs if they have submitted. The second data set is from a survey study which is conducted in 2005.[32] The survey sample is randomly selected from senior high school students who would take the CEE in 2005. This survey includes questions about socioeconomic, politic, demographic and educational background. The last data set is a list of cutoff scores for college-majors across years and publicly available on the CSSP website.

There are almost two-million students in the college admissions process but I focus on 172,951 of them in this paper. These students chose quantitative track and submitted ROLs after announcement of the college admissions scores in 2005. In addition, I restrict the focus to students who are aiming for four-year degree programs to keep my sample consistent according to admissions requirements that I described in the model section.[33]

Table 2.1 reports students' scores from college and high school entrance exams, GPAs, college choice characteristics and high school types. The number of available high school entrance exam scores is 112,055 and high school GPA is 53,041 in my data. Student high school type composition is listed in Panel C in the Table 2.1. N and S represent non-selective and selective types in the high school admissions. The average number of choices listed in ROLs is 15.63, where the maximum allowed choice is 24. Students on average submit around 9 colleges, 8 cities and 5 majors in their ROLs. Students only spare 8 percent of the ROL for the private colleges, and 19 percent for the

[31] The total score calculated as follows: $Totalscore_i = CEE_i + 0.8 \times shsGPA_i$ if the choice in the same track, $Totalscore_i = CEE_i + 0.3 \times shsGPA_i$ otherwise.

[32] For detailed information and descriptive analysis see Alkan et al. (2008).

[33] 2 year degree programs have special admission rules and are allowed to admit students without a college entrance exam score.

Table 2.1: Summary statistics for applicants 4 year degree quantitative track programs

	Mean	SD	Min	Max	Obs
Panel A. Student Characteristics					
Female	0.39	0.49	0.00	1	172951
College Score Quantitative	0.77	0.09	0.62	1	172951
College Score Verbal	0.67	0.11	0.42	1	172951
College Score Composite	0.74	0.09	0.50	1	172951
High School Score Quantitative	0.69	0.10	0.38	1	112055
High School Score Composite	0.70	0.09	0.38	1	112055
GPA	0.75	0.22	0.00	1	53041
Panel B. Choice Variables					
# Choice	15.63	7.21	1	24	172951
# College	9.42	5.14	1	24	172951
# City	8.34	5.19	1	24	172951
# Major	4.98	3.11	1	23	172951
Fraction of Private College	0.08	0.22	0	1	172951
Fraction of Evening Education	0.19	0.24	0	1	172951
Fraction of same city	0.19	0.28	0	1	172951
Mean Distance (km)	474.77	338.52	0	2500	172714
Panel C. High Schools					
Distance Educ.(N)	0.00	0.05	0	1	172951
Vocational (N)	0.03	0.18	0	1	172951
Regional (N)	0.02	0.15	0	1	172951
Vocational(S)	0.02	0.15	0	1	172951
Regional (N)	0.41	0.49	0	1	172951
Private(N)	0.01	0.11	0	1	172951
Regional(S)	0.18	0.38	0	1	172951
Anatolian(S)	0.23	0.42	0	1	172951
Anatolian Teac.(S)	0.02	0.15	0	1	172951
Private Science (S)	0.05	0.21	0	1	172951
Science (S)	0.02	0.13	0	1	172951

Notes: College entrance exam and high school entrance exam scores in Panel A are normalized according to their maximum scores to be 1. High school are presented according to their selection criterion in their admissions. (N) represents non-selective high schools and (S) represents selective high schools.

Table 2.2: Summary statistics for college in quantitative track programs

	Mean	SD	Min	Max	Obs
Cutoff score	305.20	37.23	197.95	381.18	1752
	(68.59)	(26.18)	(0.01)	(99.92)	
Max entrant score	322.47	29.89	213.11	399.03	1752
Evening Educ.	0.15	0.36	0	1	1879
Private	0.27	0.44	0	1	1879
Full Sch.	0.03	0.16	0	1	1879
Partial Sch.	0.08	0.27	0	1	1879

Notes: Private colleges offer mainly two types of scholarship opportunities. Full scholarship covers all tuition expenses and includes fellowships in some cases. Partial scholarships only cover a fraction of tuition, the remaining part has to be paid from students. The numbers in parentheses presents admission percentiles.

evening education[34] and colleges in the same cities.

Table 2.3: Summary statistics for college cutoff scores by college types

College type	Threshold type	Mean	SD	Obs
Public	Cutoff score	308.45 (69.49)	32.12 (24.19)	1338
	Max. entrant score	323.17	27.48	
Private w/o scholarship	Cutoff score	271.40 (48.68)	43.98 (32.85)	251
	Max. entrant score	307.35	36.30	
Private w. scholarship	Cutoff score	330.58 (83.53)	31.17 (20.66)	163
	Max. entrant score	339.99	26.85	

Notes: Private w/o scholarship represents all programs that are offered from private colleges and do not offer any funding opportunity. Admitted students are required to pay their tuition. The numbers in parentheses presents admission percentiles.

Table 2.2 reports college characteristic and threshold type summaries. There are mainly two types of colleges in the Turkey, public and private. Private college attendance requires high tuition payments, while public colleges are mostly funded by the state.[35] In the quantitative track, there are 1,879 4-year degree programs and 27 percent of these programs are offered by private colleges. Private colleges offer scholarship opportunities which are awarded based on college admissions outcomes. There are 11 and 30 percent of prog- rams are offered by private colleges provide full or partial funding, respectively.

[34]Evening education is a program that offers education at evening time. These programs admit students as in the same admission procedure but they have different quotas in the system. Hence, a college can offer both normal program and evening education program for the same major in the same college admission period.

[35]Except for some administrative fees, there is no tuition at these colleges.

Table 2.4: Summary statistics for college cutoff scores by selected college-majors

Majors	Threshold type	Mean	SD	Obs
Medical School	Cutoff score	352.98 (96.32)	7.90 (2.54)	53
	Max. entrant score	361.86	9.17	
Electronic Eng.	Cutoff score	327.50 (82.73)	34.01 (23.04)	76
	Max. entrant score	342.93	21.46	
Computer Eng.	Cutoff score	326.05 (81.97)	38.13 (22.49)	82
	Max. entrant score	345.09	20.36	
Industrial Eng.	Cutoff score	323.84 (81.22)	37.24 (24.35)	62
	Max. entrant score	340.65	23.07	
Mechanical Eng.	Cutoff score	318.59 (80.02)	23.83 (14.38)	79
	Max. entrant score	330.01	18.21	
Civil Eng.	Cutoff score	315.65 (76.42)	32.93 (15.56)	75
	Max. entrant score	330.32	24.95	
Math	Cutoff score	306.86 (66.88)	32.45 (30.02)	132
	Max. entrant score	319.27	27.76	

Notes: The numbers in parentheses presents admission percentiles.

More detailed summary statistics for private and public college cutoffs are presented in Table 2.3. There are two types of thresholds are presented, cutoff score and maximum entrant score. These thresholds represent minimum and maximum scores of admitted students in the realized matchings. According to 2004's matching outcome, cutoff scores are differentiated according to funding and scholarship types. Private colleges have high thresholds if they provide scholarship opportunities while programs that do not offer funding have lower thresholds.

Table 2.4 reports two threshold types from 2004's matching by selected competitive majors in the quantitative track. The mean cutoff scores show that there is a high degree of stratification among majors according to admission levels. Medical school is the highest in terms of average cutoff scores.

Another important college choice factors can be student location and distance preferences. In many countries colleges are not distributed evenly over the locations. In Turkey, majority of colleges are located in the crowded cities.[36] Moreover, there are some regions that do not have any close distance colleges. Heterogeneity in terms of college locations may have impact on students' college choices and college admissions preparation strategies. In table 2.5, I present college

[36]See Figure A.7 for the number of colleges by cities in Turkey.

threshold differences between 3 biggest cities of Turkey; Istanbul, Ankara, Izmir and the remaining cities. The mean cutoff scores show that colleges in 3 biggest cities have higher cutoff scores than the colleges in the other cities.

Table 2.5: Summary statistics for college cutoff scores by college locations

Locations	Threshold type	Mean	SD	Obs
Istanbul, Ankara, Izmir	Cutoff score	323.19 (78.60)	33.69 (24.10)	635
	Max. entrant score	341.20	23.66	
Other cities	Cutoff score	294.98 (62.59)	35.22 (25.59)	1117
	Max. entrant score	311.82	27.76	
Total	Cutoff score	305.20 (68.59)	37.23 (26.18)	1752
	Max. entrant score	322.47	29.89	

Notes: Istanbul, Ankara, and Izmir are the biggest 3 cities of Turkey and they are hosted majority of colleges in the country. The numbers in parentheses presents admission percentiles.

Finally, to quantify student time and effort investments for the exam preparations, I use three measures that show student study times by themselves and amount of time is spent in private college exam prep-courses and tutoring. These measures are obtained through survey questions. Table 2.6 presents preliminary estimation results of student time and effort investments in the college entrance exam preparations according to their college choices two different competitiveness level contest sets. According to regression results, students who choose private colleges more in their ROLs have lower self-study and prep-course times, on the other hand students who choose medical schools more spend higher time for self-study and prep-courses. For private tutoring, the results are opposite.[37] Based on these results, I can claim that there is a significant different in time and effort investments for college entrance exams in terms of students' college preferences.

Cutoff score differences across college types, majors and locations are presented in the Tables 2.3, 2.4, and 2.5. In order to see optimization differences in college admissions process, I need to apply the econometric techniques that have been proposed in the Section 4.

[37] Private tutoring option is only available for higher income people. Positive and significant higher private tutoring result is not surprising for students who have sufficient income to go private colleges.

Table 2.6: Preliminary estimation results of student effort and time investments for college entrance exam preparations

	Not Competitive	Competitive
Variables	Private College	Medical School
Prep Course	-0.010	0.008
	(0.005)	(0.002)
Tutor	0.079	-0.016
	(0.004)	(0.002)
Self Study	-0.041	0.017
	(0.005)	(0.003)
Constant	0.243	-0.011
	(0.023)	(0.010)
R^2	0.14	0.03
N	4305	4305

Regression results are obtained from OLS estimations. Private colleges and Medical school fractions are constructed as it is described in the college choice form contest sets.

2.5.2 **Results**

This section reports the estimates for the parameters of interest. Since my method requires focusing on one dimension of students' college preferences vector, I investigate the effects of preferences for private college, major, and location separately.

2.5.2.1 **Private colleges**

The first model in table 2.7 presents private college choice estimation results when endogeneity is not taken into account. The (2) and (3) models report CF and QML estimation results that are described in section 4. These two models' results make evident that taking preparation endogeneity into account in the estimation creates difference in the interpretation of student preferences for colleges. In particular, the college exam scores effects on private college choice not only changed in terms of magnitude but also reversed in sign.

According to private college choice estimation results under exogeneity assumption in the model (1), I infer that students who have lower scores more likely to choose private colleges. On the contrary, the model (2) and (3) shows that students that have higher scores more likely to choose private colleges. In addition, there is a significantly negative relationship in the college entrance score formation equation's and private college choice equation's residuals. Note that residuals from the college exam score formation equation and private college choice equation include an unobserved preference factor in the econometric framework. Hence, the correlation of the residuals which is reported with ω and ρ represents the relationship between the college choice from different contest sets and score formation.[38] The results in the model (2) and (3) suggest that a student who has higher inclination to choose private colleges in the college admissions has lower incentive to obtain higher score on the college entrance exam. This relationship is result of the difference in the competitiveness between private and public colleges. Recall from the Table 2.3, private colleges without scholarships have lower cutoff scores than public and private colleges with scholarship programs. As I show in the college contest set framework, students who prefer colleges with low thresholds have less incentive to obtain higher entrance exam score. In models (4) and (5) college contest choice indicators are replaced with unweighted fractions and binary variable which takes one if fraction of private colleges are more than 50 percent and the first choice in ROL is private college. The results are aligned with the findings in the model (2) and (3).

In order to control for any exam day shock effects, I alternatively estimate preference for private college effect on college admissions preparation strategies with the scaled high school GPA (*shsGPA*) score. *shsGPA* is the other score in the admission criterion and has a considerable effect on it. GPA is an average of many courses from high school, consequently it is less prone to a single exogenous shock in the formation than college entrance exam score. In addition, since GPA is scaled according to the average of students' successes from the same high school in the college entrance exam, score normalization according to difficulty level between high schools comes immediately. The model (6) in table 2.7 reports estimation results with the *shsGPA* variable instead of college entrance exam score. The results are aligned with the findings in the model (3). The main difference is in the gender effect on score formation. Female students have better performance in

[38]Since the correlations under CF and QMLE is qualitatively the same and have very close magnitudes, I prefer to continue to use QMLE results for interpreting the results. The correlation coefficient is simpler way show the direction and magnitude of the relationship. ω estimate in CF equals ρ/τ in the QMLE.

Table 2.7: Estimation results for private college choice

	(1)(Frac-QMLE)		(2)(Frac-CF)		(3)(Frac-QMLE)		(4)(Frac-QMLE)		(5)(Bin.-QMLE)		(6)(Frac-QMLE)		(7)(Frac-QMLE)	
Model	Exo.		Endo.		Endo.		Endo.		Endo.		Endo.		Endo.	
Variables	Coef	St Err	Coef	St Err	Coef	St Err	Coef	St Err	Coef	St Err	Coef	St Err	Coef	St Err
College Entrance Score	-2.33	0.06	1.75	0.12	1.63	0.20	1.65	0.20	1.12	0.20	·	·	-0.67	0.36
Scaled high school GPA	·	·	·	·	·	·	·	·	·	·	1.36	0.10	·	·
Female	-0.21	0.01	-0.13	0.01	-0.12	0.02	-0.12	0.02	-0.13	0.02	-0.29	0.02	-0.14	0.05
Constant	0.65	0.15	-2.72	0.16	-2.54	0.32	-2.58	0.33	-2.12	0.24	-1.29	0.20	-4.82	0.52
High School Dummy	Yes		Yes		Yes		Yes		Yes		Yes		Yes	
City Dummy	Yes		Yes		Yes		Yes		Yes		Yes		No	
Father Educ.	No		No		No		No		No		No		Yes	
Mother Educ.	No		No		No		No		No		No		Yes	
Income	No		No		No		No		No		No		Yes	
Wealth	No		No		No		No		No		No		Yes	
ω	·	·	-5.04	0.15	·	·	·	·	·	·	·	·	·	·
ρ	·	·	·	·	-0.34	0.03	-0.35	0.02	-0.41	0.02	-0.45	0.02	-0.10	0.04
τ	·	·	·	·	0.07	0.00	0.07	0.00	0.07	0.00	0.04	0.00	0.10	0.01
Obs	110,807		110,807		110,807		110,807		110,807		45,722		3,507	

Notes: The model (1) reports results with exogenous college entrance exam scores. In the models (2), (3), (4), (5), (6) and (7) college entrance exam score and scaled high school GPA are endogenous. High school entrance exam scores are used as exogenous variation in the CF and QMLE models. Standard errors are calculated from bootstrap simulations with 100 replications for these models. For detailed presentation of estimation result of the models, I refer reader to the tables A.27, A.28, A.29 and A.35, in Appendix D.

high school GPA formation, while the reverse is true for the college entrance exam scores.[39]

In the all models, I include high school types dummies. High school types do not only differ in terms of student selection mechanisms. Out of pocket expenses of private high schools can be considerable amount in the family budget.[40] Estimation results show that students from private high schools tend to choose private colleges more than students from public high schools. These results suggest that socioeconomic status have significant effects on college type choices.

The model (7) in table 2.7 reports QMLE results obtained from the survey sample. The correlation coefficient between choice and score formation residuals is still significant, while its magnitude is smaller. This specification shows that even financial factors are controlled which can be considered the most important criteria in the private college choice, score formation strategies and college choice are still significantly related. This result is a substantial finding for the preference effect on the score formation strategies.

2.5.2.2 **Majors**

The competitiveness differences across college majors is presented in Table 2.4. This section presents student behavior analysis under this information. Table 2.8 presents estimation results when the endogeneity of score formation is not taken into account. Table 2.9 reports estimation results of QMLE estimation results for selected major preferences with endogenous entrance exam score.

In Table 2.8, the major choice is significantly affected from the entrance exam scores of college applicants. Medical schools are selected by higher scored students in comparison to other majors. In addition, gender plays a significant role in major choice after scores are controlled for. Female students prefer medical school and mathematics departments, while male students prefer engineering departments.[41]

[39]Niederle and Vesterlund (2007) and Buser et al. (2014) analyze competitiveness difference between female and male students. They find that women shy away from competition. This may be another reason behind the difference in male and female students performance apart from their differences in preferences for colleges.

[40]In 2005, the average cost of attending a private high school in Ankara (the capital of Turkey) corresponds 14

Table 2.8: QMLE result for major choice without endogenous college entrance exam score

	Medical School		Engineering		Math	
Variables	Coef	St Err	Coef	St Err	Coef	St Err
College Exam Score	14.31	0.22	1.39	0.04	4.17	0.05
Female	0.23	0.01	-0.70	0.01	0.10	0.01
Constant	-14.91	0.27	-0.97	0.10	-5.26	0.11
High School FE	Yes		Yes		Yes	
Obs	110,807		110,807		110,807	

Notes: Estimation results with full set of regressors is presented in Appendix D in Table A.30.

Table 2.9: QMLE results for major choice with endogenous college entrance exam score

	Medical School		Engineering		Math	
Variables	Coef	St Err	Coef	St Err	Coef	St Err
College Exam Score	13.21	0.24	3.26	0.09	0.09	0.14
Female	0.18	0.01	-0.65	0.01	0.02	0.01
Constant	-14.10	0.30	-2.45	0.11	-1.88	0.16
High School FE	Yes		Yes		Yes	
ρ	0.11	0.01	-0.17	0.01	0.35	0.01
Obs	110,807		110,807		110,807	

Notes: Standard errors are calculated from bootstrap simulations with 100 replications. Estimation results with full set of regressors is presented in Appendix D in Table A.31.

According to table 2.9, major choice and college entrance exam score formation are significantly related. In addition to students who have higher scores on the college exam more likely to choose medical schools, there is a positive correlation between medical school choice and college entrance exam score formation equations' residuals. This result indicates that students who has higher inclination to choose medical school put more effort into college score formation than the others. A similar case is also observed in mathematics major choices. On the other hand, the results for the engineering departments are not representing the similar pattern. Even though students who have higher scores choose engineering departments, there is a negative correlation between the choice and score formation equations' residuals.[42]

months of minimum monthly wage.

[41] Gender differences in major choice is documented in many other college-major choice settings in the literature. Among many others, Wiswall and Zafar (2014) and Zafar (2013) show gender gaps in major choice in the US.

[42] Table A.32 reports the average cutoff percentiles of engineering majors by private and public colleges. The difference in the competitiveness levels in the private and public colleges may be the one of the the underlying reasons for this result. However, Medical schools departments are predominantly offered by public colleges. Therefore, the competitiveness differences of these majors are less diverse.

2.5.2.3 **College locations**

In order to understand location and distance preferences on college admissions preparation strategies, students' location choices from their submitted ROLs are analyzed in this section. Table 2.10 presents estimation results of fraction of the same city choices in their ROLs under 3 different specifications. I restrict the sample with the students from cities that are not from the biggest three cities and their neighbors. By this way, I can analyze students' same city preferences where the competition for these colleges are differentiated from the colleges in the biggest cities.

Table 2.10: QMLE results for the same city choice under different specifications

	(1)		(2)		(3)	
	(Frac-QMLE)		(Frac-QMLE)		(Frac-QMLE)	
Model	Exo.		Endo.		Endo.	
Variables	Coef	St Err	Coef	St Err	Coef	St Err
College Exam Score	-0.67	0.06	1.06	0.38	·	·
Scaled HS GPA	·	·	·	·	0.49	0.39
Female	0.13	0.01	0.17	0.02	0.11	0.05
Constant	0.03	0.17	-1.37	0.43	-0.49	0.55
High School FE	Yes		Yes		Yes	
ρ			-0.15	0.03	-0.16	0.08
Obs	37150		37150		13722	

Notes: I only consider students from cities that are not from the biggest 3 cities and their neighbor cities. Standard errors are calculated from bootstrap simulations with 100 replications for models (2) and (3).

According to first specification that presents fractional response QMLE results without endogenous college entrance exam score, students who obtain lower scores from the college entrance exam choose colleges in the same cities. On the other hand, the second model that presents QMLE results with endogenous college entrance exam score show the opposite case. There is also an analogous result in the model (3) where scaled high school GPA replace college entrance score. In addition, ρ is significant and negative in sign in both model (2) and (3) which indicates that same city preferences for small cities decrease motivation for obtaining higher scores from college entrance exam.

Another interesting result from table 2.10 is the heterogeneity between male and female students' same city choices. Female students are choosing colleges in the same cities significantly more than male students. This result is another basic evidence for the heterogeneity in college choice.

2.5.3 Variance decomposition in college entrance exam score formation

The significant effects of college type, major and location preferences on college exam score formation strategies have been shown in the previous subsections. In this subsection, I decompose the variation of the college entrance exam scores to quantify the overall effect of preferences in the score formation strategies.

Measuring variation of the preference effect in the score formation requires a special setup because these effects are part of the unobservables in the college admissions criteria's equations. Consider a college entrance admissions criterion that is the average of at least two scores;

$$\tilde{p}_t = h\gamma + \eta + e_t \text{ where } t = 1, \ldots, T \geq 2$$

where h denotes high school entrance exam scores. η denotes unobservable preference factor which includes all preference components. e_t denotes exogenous shock in the admissions criteria scores. Assume that $e \perp \eta, h$ and $e_i \perp e_j$ where $i \neq j$. In addition, $(\tilde{e})$ has normal distribution, where $\tilde{e} = e + \eta$. If there is an consistent estimator for γ[43], I can estimate the variance of η using the residuals of admissions criteria scores.

$$\Sigma = \begin{pmatrix} var(\tilde{e}_1) & cov(\tilde{e}_1, \tilde{e}_2) \\ cov(\tilde{e}_1, \tilde{e}_2) & var(\tilde{e}_2) \end{pmatrix} = \begin{pmatrix} var(e_1) + var(\eta) & var(\eta) \\ var(\eta) & var(e_2) + var(\eta) \end{pmatrix}$$

where Σ denotes variance-covariance matrix of $(\tilde{e})$. Since these matrix can be estimated by the residuals from the admissions criteria scores, I reach variances of η, e_1, e_2.[44]

The results show that in the college entrance exam scores; 16%, 41% of the variations come

[43] I use the most conservative estimates for γ from the econometrics specifications in the subsections above.

[44] The distribution of η can be estimated with non-parametric method developed by Li and Vuong (1998). This method is applied in auction setting Li et al. (2000) and earning dynamics Bonhomme and Robin (2010).

from preference heterogeneity and high school entrance score respectively. The majority of the variation comes from other independent factors that may include the exam day shock consists of 43%. For the scaled high school GPA score; preference, high school entrance score, and other factors variations are 41%, 54% and 5% respectively. These results shows that preferences are an important part of the college admission criteria score formation strategies. In addition, there are considerable amount other independent factors' effect on college entrance exam scores.[45]

2.5.4 Discussion: Affirmative action policies with endogenous entrance scores

I discuss the college admissions model to explore counterfactual questions of interest in the affirmative ac- tion policies for the college admissions. First, how do affirmative action policies affect student admissions preparation strategies under the same preference and competition structure? Second, how should college admissions policies be designed to take preference effects on admission scores into account?

I first explore how affirmative action (AA) policies affect students' preparation strategies. The main rationale behind the affirmative action policies is creating incentives to increase admissions into selective colleges for students from disadvantaged socioeconomic backgrounds or minorities. In my setting, these policies can be considered as cost reduction tools of obtaining score by providing preferential treatment in college entrance score formation. In other words, a student i obtain a score $\tilde{p}_i$ but in the matching stage her score is treated as $d\tilde{p}_i$ where $d > 1$ under an affirmative action policy.[46] Thus $\tilde{p}$ replaced with $\tilde{p}/d$ in the cost function in equation 2.4 for the treated student's optimization. The consequence of AA policies, a treated student's strategy for college admissions exam preparation can be affected in two ways. First, the treated student finds the same colleges optimal but because of new cost structure, she does not need to put the same effort as before. Sec-

[45] In order to decompose the variation of admission scores and preferences in the submission of ROLs, I run naive regressions from selected choice variables that are constructed from ROLs. Table A.33 reports the explained variations of the selected choice variables by admission scores, gender, high school type and city dummies.

[46] Bodoh-Creed and Hickman (2016) show that under large college admission contests, two common families of affirmative action policies, admission preference and quota, have the same sets of equilibria. Using this result, I do not investigate different AA policies.

ond, under new cost of obtaining structure the treated student starts to find different colleges from another contest sets are favored and changes her preparation strategy accordingly. Even though both strategies increase expected utility of the treated students, their matching outcomes can be fairly different. The existence of the first case is the main reason for the reduction in the admission score based AA policies' effectiveness that is aiming to increase a group of students admissions for targeted colleges, where these colleges are not preferred from some of the students from the designated group.

Given that student valuations for colleges can substantially affect college entrance exam preparation strategies, I next turn to explore the preference intervention effects on college admission process. The importance of preference intervention for different socioeconomic groups is to introduce more information about college and major effects on future labor market outcomes as well as lifetime earnings. This intervention can be considered as information provision about the college and majors. As it is observed from college and major choices in Turkey, there is significant variation in choice patterns of different groups of the society.[47] The choice heterogeneity among different groups in the society is not only originated from preference differences but also informational difference about the opportunities of selected education paths[48] as well as other choice preferences such as risk. Disparities in college and major choice differences between children from low- and high-income families and genders underscore the importance of policies that raise the selection and preparation issues among disadvantaged students. Therefore, influencing preferences or designing college supply by taking students' preferences into account can be applied as a new type of admission policies to improve outcomes of students from disadvantaged socioeconomic and information backgrounds.

Specifically, the estimation results from the empirical application points the heterogeneity between male and female which is a notable topic in the affirmative action policies in the college

[47] Major choice difference and college location differences are considerably different between male and female students. In addition, heterogeneity in major choice from different socioeconomic backgrounds indicates the preference heterogeneity among various social groups.

[48] Among others, Saygin (2013) and Caner and Okten (2010) investigate gender and socioeconomic factor effects on major choices in Turkey.

admissions. Female students choose the colleges from their hometowns more likely, while choose private colleges less likely. Moreover, female and male students major choices are different than each others on average. I also observe that female students' college entrance exam scores are lower than males on average, while high school GPA is the opposite case.[49] Since student abilities are controlled with the exogenous test scores in the empirical application, divergences in college exam results and choices are expected to be related with different preferences between male and female students. These findings suggest that preference differences have substantial effects in the college entrance exam preparation strategies.

To fix ideas, let's consider an hypothetical affirmative action policy example related with the female students' location choices. Female students tend to choose college in their hometowns more in Turkey. The policymaker wants to increase gender diversity from different regions in the selective colleges and consider an admission preference rule[50] for female students in the distant locations and only feasible for colleges that are away from their hometowns. If a competent female student who is far from selective college cities prefers to be assigned with a college in her hometown, the proposed AA policy does not affect her college admissions' strategy if the treatment does not lower her cost sufficiently. Under these conditions, admission preference policies that aim to redesign college admissions process are not effective to increase diversity in selective colleges if the majority of female students choose colleges in their hometowns because of the preference reasons. On the other hand, an alternative affirmative action policy that increases female students' college valuations for targeted by providing opportunity information, better college environment that decreases concern of families or financial aid incentives may influence female students college admissions process strategies and increases the likelihood of targeted admissions outcomes.

As it can been seen from the example above, intervening college admissions process by providing preferential treatment on student priorities does not lead targeted results directly for the admissions outcomes. Heterogeneity in college preferences among candidates changes the effec-

[49]Summary statistics of college applicants by gender can be found in the table A.21.

[50]Admission preference rule is a markup function and students' score are transformed according to markup. In other words, students who are benefiting from the policy are repositioned according to markup function.

tiveness of score oriented AA policies since there is no general consensus of college preferences among students. Furthermore, distortions which are the results of AA policies would lead worse outcomes for the untreated students. Consequently, policies that do not take the preference effects on score formations into account may cause welfare losses rather than targeted improvements, which is an open question to be answered.

2.6 **Conclusion**

College admissions have two important stages; preparations for the admissions and the matching between students and colleges. The analysis of college admissions have been focused on only one part of the process under simplification assumptions for the other part. In this paper, I analyze college admissions preparations and student-college matching stages together with the help of large matching market model. My college-student matching model captures student behavior where they endogenously choose their priorities for the college admissions by obtaining a college entrance exam score. Under the continuum of students assumption, the matching outcome is approximated with a distribution of college cutoff scores. Using this result, I model student's behavior in the college admissions preparation stage as an individual optimization problem over risky investment projects.

Characterization of the college admissions process as an individual investment problem allows me to analyze student preferences effects on college admissions preparations. To capture student preference information from the submitted ROLs, I introduce college contest sets. In this framework, colleges are partitioned into mutually exclusive sets according to their observable characteristics. If there is difference in the aggregate competitiveness between the contest sets, I can identify the effects of students' college preferences on the entrance exam preparations because getting to be admitted to colleges in different contest sets can be done with different strategies.

In the econometric framework, I develop a simple econometric method to analyze the effect of students' college preferences on college exam performance. With this method, I analyze college admissions in Turkey in the empirical application. The estimation results show that there are significant relationships between college choices from different contest sets and college entrance exam preparation strategies under different specifications. The differences not only change the interpretation in terms of magnitude of the effect of score variables on choice, but also change the direction of the effect. Consequently, disregarding endogenous admission preparation strategies in the college admissions and public school choice models to estimate students' preferences is not innocuous.

The significance of the relationship between choice and score formation suggests the evaluation of affirmative action policies under the endogenous score formation framework. College admissions data from Turkey show that there are variations in college choices from different social groups. It has been shown that students' college preferences change their college admissions preparation strategies, therefore affirmative action policies that are only designed to influence students' behaviors with the admission score policies lose their effectiveness. This finding supports researches (among others, Hoxby and Avery (2013), Goodman (2016)) that investigate the difference in college choice of high-achieving low-income students from high-income counterparts.

A further structural analysis of the college admissions data to estimate students' college preferences from rank order lists taking endogenous score formation into account would be a natural direction of future research. This extension would allow for a valid prediction for students' college preference estimations rather than depending on results that only from the truncated rank order lists. This is also important in policy design perspective such that estimation results provide the relationship between college admissions preparation and college choice. Thus, a policymaker can design admission policies using preference information rather than only restricting itself with admissions criteria as the only policy tool.

CHAPTER 3

Exploring College major Choices: Does Marriage Matter?

with Tong Li

3.1 Introduction

Considering education as a premarital investment in addition to human capital investment is not a new phenomenon in the economics literature after the seminal paper Becker (1973). The relationship between marriage market and education decisions are investigated in many papers (e.g., Iyigun and Walsh (2007), Chiappori et al. (2009), Goussé et al. (2017), among others) and the effects of marriage expectations on students' college participation decisions are quantified empirically for young women (Ge (2011)). To this end, one can consider the effects of marriage expectations on decisions in college education (e.g., major and minor choices, duration of higher education, activities during the college period) because subsequent household - labor market time allocation and expected lifestyles are strongly associated with these choices. In this context, it is also important to examine the effects of marriage expectations on decisions in college education to design better labor market, education, and marriage interrelated policies.

In this paper, we empirically investigate the effects of marriage expectations on college major choices. Differences in marriage expectations have potential impacts on college major choice because college majors differ in many aspects including expected lifestyle and occupations. Students may consider expected "work-life-family" time allocations when they are choosing majors. Moreover, data drawn from the 2013 survey year of the National Longitudinal Survey of Youth 1997 (NLSY97) cohorts document differences in marriage realizations of college graduates from different majors.[1] Selection of a major with different marriage expectations can present one potential explanation for why there are observed differences in the marriage outcomes. Therefore, an

[1]At the time of the survey, the mean marriage rate of Education major's graduates exceeded the overall average rate (57%) by more than 10%, while Computer and Information Science major graduates' mean marriage rate was 10% lower than the overall average rate.

empirical procedure is required to understand motivational differences of college students' major choices and test the effects of marriage expectations on major choices. In addition, understanding motivational differences of college major choice is an important part of the talent allocation within the economy and composition of skills in the labor market. In particular, apart from the pecuniary incentives as a solution to shortages in various skills in the market, marriage friendly benefits may lead to an efficient allocation of skills.

We propose a copula-based econometric framework that incorporates multinomial college major choices into marriage realizations to test the effects of marriage market expectations on college major choices. Since marriage expectations are unobservable in our framework, we adopt a selection-on-unobservables model to test their effects from data. Econometric models with selectivity have been also used in self selection models with multinomial choice cases (e.g., Lee (1983) and Dahl (2002)), where outcome variables are continuous. This paper develops a self selection model with a binary outcome variable, which is marriage realization in our application. The proposed approach could be easily applied in many cases where the outcome variable is a binary response and there are multiple treatments affect the outcome variable.

The existing empirical literature on education decisions generally considers the effects of expected earnings. The seminal paper by Willis and Rosen (1979) shows the positive impact of the expected earnings on college education under the rational expectation assumption. Keane and Wolpin (1997) and following papers Eckstein and Wolpin (1999), Belzil and Hansen (2002) estimate structural dynamic models of schooling decisions and find significant positive impacts of expected earnings on college participation. More recently, Arcidiacono (2004) and (2005) focus on the effect of expected earnings on major choice with sequential models of college attendance, accounting for both the demand and supply sides of schooling.[2] As a consequence, it is important to control expected earnings' effects on college major choice to test the effects of marriage expectations.

More recent studies on college major choice focus on non-pecuniary motivations. Beffy et al.

[2]Kirkeboen et al. (2016) investigate the payoffs to different types of postsecondary education in a special identification framework benefiting from Norwegian education system.

(2012) examine college major choice and find a small but positive significant effect of expected earnings in the context of French universities. Their results show that non-pecuniary factors are the main determinants. Wiswall and Zafar (2015) study the determinants of college major choice using experimentally generated beliefs and find that heterogeneous tastes are the dominant factors in the choice. Based on recent findings of the effects of heterogeneous tastes and the anticipated impact of the marriage expectations on major choice, an empirical procedure can be proposed to test the significance of the relationship between major choice and marriage outcomes. To the best of our knowledge, this is one of the first papers in the literature that investigate the relationship between college major choice and marriage expectations using individuals' college major choices and marriage realizations.[3]

The papers that focus on female labor market outcomes have also related results to our paper. Keane and Wolpin (2010) examine women's sequential school attendance, work, marriage, fertility, and welfare-program participation decisions. They show that marriage market opportunities change schooling decisions. Flabbi and Moro (2012) analyze the job flexibility effects on female labor market outcomes and they show that wage and schooling differences between male and female may be related to flexibility preferences. Bronson (2014) studies the gender differences in major choices and she finds that women prefer degrees which provide more "work-family flexibility". The findings from these papers suggest that differentiation in the "work-life-family" balance preferences between female and males may be originated from differences in their marriage expectations (or marriage related preferences) which may also affect their major choices.

Although we focus on US college graduates in this paper, the same method can be readily applied to data from other countries. Currently, 58% of college graduates are female and the proportions of the female students across majors vary much.[4] Figure A.8 presents the historical changes in the college participation and proportion of degrees awarded to female across majors

[3]Bicakova and Jurajda (2016) and Bičáková and Jurajda (2017) are the two closest papers that investigate the relationship between college majors and marriage, however, their focuses are gender composition in majors, early fertility, and homogamy within the majors.

[4]Note that the proportion of full time workers is lower for female graduates. This may be another sign of motivational differences in college attendance and college major choice. For detailed information, see the National Survey of College Graduates.

in the US. Education and Health related majors are dominated by females, while Science, Engineering and Business are male-dominated majors. Motivated by these differences, we also analyze the gender differences in the effect of marriage expectations on major choices. This is a natural direction to investigate because male and female are the two sides of the traditional marriage. We apply our method to the data provided from the National Longitudinal Study of Youth. This dataset includes detailed information of participants' demographic backgrounds, marriage realizations as well as their high school and college related variables. After students' characteristics are controlled for, we find that graduates from Education major marry more. Females marry earlier, while male marry later. In order to investigate the effect of marriage expectations on college major choices, we compare average treatment effects of chosen majors on marriage realizations with and without selection-on-unobservable specifications. Our test results show that the effects of marriage expectations cannot be rejected, even after individual characteristics and expected earnings are controlled for. Moreover, these findings are changing between male and female students.

This paper is organized as follows. Section 2 describes the econometric framework for college major choice and marriage realizations. The testing procedure and specification of the model are discussed in Section 3. Section 4 presents the data and summary statistics. In Section 5, we present preliminary estimation results and then apply our proposed test. Finally, Section 6 concludes.

3.2 **Econometric Framework**

After graduating from high school, students, who continue to college, are choosing majors to specialize. There are many determinants in college major choice. The characteristics of the chosen major are important factors for shaping graduates' future life outcomes, and therefore they are expected to be considered in the choice. In our paper, we focus on the effects of marriage expectations on college major choice. Our model allows that each student has personal marriage expectations in addition to the other personal heterogeneities. The personal marriage expectations, which obviously affect the marriage decisions and realizations of college graduates, also enter into the college major decision problem. Then, the econometric problem turns to test the significance of the effects of marriage expectations on college-major choices from observed college majors choices and marriage realizations.

The main challenge for identification is the possibility of the other unobservable factors' simultaneous effects on marriage realizations and college major choice. Even though marriage expectations is the underlying component that is considered in this paper, there are other potential factors, (e.g., ability, pecuniary motivations, and labor market preferences) may have effects on both college major choice and marriage realization. In our econometric specification, we control these variations either using well accepted measures[5] or adopting standard assumptions[6] which help us to construct control variables from the data.

3.2.1 Major choice

In the US higher education system, each college student must choose an area of concentration during the early stages of undergraduate education.[7] We assume that there are J majors, which is indexed with j, and a college student must choose one of them. In the model, we consider that major choice depends on the student's expectations about graduating from the corresponding major. These expectations include labor market outcomes, marriage market outcomes, and the other taste variables. In order to keep the model simple, we only consider undergraduate education level and exclude decisions for continuation to higher degrees.[8]

Student i derives V_{ij} utility from choosing major d_j $(j = 1,\ldots,J)$. We assume this utility is a composition of two elements, namely, pecuniary value and non-pecuniary value, denoted as v_{ij}^p and v_{ij}^n, respectively. Pecuniary value, v_{ij}^p, is the expected labor market earnings conditional on graduating from major d_j. We normalize the effect of the pecuniary motivation to 1. Pecuniary value of major d_j can be written as

[5] In our estimations, we use first three term of college GPAs of students to capture ability effects.

[6] With the rational expectation assumption, we construct potential wage and working conditions of selected majors.

[7] Major selection process differs in many countries in terms of timing. In some countries, students are accepted to higher education institutions with their major choices in the admissions process.

[8] There is no duration of education decision made by students. This restriction is compatible with our data since less than 5 percent of students have graduate degrees. Extension with the further degrees can be easily applied to our testing procedure.

$$v_{ij}^{p} = E(w_{ij}) \tag{3.1}$$

where w_{ij} denotes the expected average earning associated with major j for student i. By following the empirical literature, we adopt the rational expectation assumption for characterizing students' expected earnings. In other words, students are assumed to know their expected earnings for graduating from each major choice given the labor market conditions.[9] Therefore, we can allow earning heterogeneities with a simple and widely accepted methodology in the literature.

The second part of the utility consists of all other components apart from the pecuniary value associated with major d_j. In other words, it includes student's expectations about future lifestyles, tastes for the majors, and their marriage expectations. Therefore, we can write non-pecuniary value of major j for individual i in a simple additive form as

$$v_{ij}^{n} = z_i\beta_j + u_{ij} \tag{3.2}$$

where z_i represents student i's characteristics (including gender, school performance variables, race, and regional characteristics), β_j is the corresponding parameter vector associated with the major d_j. u_{ij} is the unobservable part of the equation and it includes all unobservable non-pecuniary factors such that marriage expectations, leisure preferences, so and so forth.

Student i chooses major d_{ij^*} which gives the highest expected utility.

$$d_{ij^*} = arg\ max_{j\in\{1,\ldots,J\}} V_{ij} \tag{3.3}$$

where $V_{ij} = v_{ij}^{p} + v_{ij}^{n}$.

[9] In our estimation, we use average earnings to estimate counterfactual the expected earnings for other majors. Under the assumption of observable characteristics are independent of unobserved ability, we can predict expected earnings in different majors based on our estimation. In this way, we simply assume that unobserved ability affects expected earnings in different majors proportionally.

3.2.2 **Marriage**

We assume that individuals have marriage expectations and these affect their marriage outcomes. Specifically, if an individual has higher marriage expectations, one can expect to see that individual has the higher chance of getting married.

It is expected that marriage outcomes are combination of many other factors, which are not restricted with the person-specific marriage expectations. We prefer to write a parsimonious marriage outcome equation, instead of modeling it explicitly.[10] Therefore, we write college graduate i's marriage outcome equation as

$$m_i = 1(x_i\gamma + \sum_{j=1}^{J} d_{ij}\tau_j + \varepsilon_i > 0) \tag{3.4}$$

where m_i is a marriage dummy which takes one if the individual is get married within the sample period. Marriage dummies capture realized outcomes, and therefore our results can be explained as revealed preferences rather than stated preferences.

We control observable individual characteristics (age, racial and regional characteristics, ...), which is indicated by z_i. d_{ij} is an indicator for individual i's major choice j. It becomes 1 if the individual i chooses major d_{ij} in her college education. By construction at least one of the major dummies has to be one and it must be the only nonzero dummy. These dummies control the effects of a chosen major on marriage outcomes. They also capture the dynamic effects of chosen majors. In other words, they include major specific effects in the marriage market.[11] γ and τ_{ij} are unknown parameters to be estimated. ε_i represents the unobservable part of the equation 3.4. It includes personal marriage expectations as we explained above and idiosyncratic terms in marriage outcomes.

[10]The two-sided nature of marriage directs us to choose a simple structure rather than a complicated model. In addition, partner search models that are used for marriage models have similar structure and assumptions.

[11]These dummies control differentiated demands in the marriage market in terms of majors (occupations). For example, if there is a higher demand for Education major graduates to get married, the Education major dummy absorbs this effect.

3.3 Setup for Estimation

3.3.1 Model specification

According to our econometric framework, the binary marriage outcome and the college-major choice equations take the following forms

$$m_i = 1(x_i\gamma + \sum_{j=1}^{J} d_{ij}\tau_j + \varepsilon_i > 0) \tag{3.5}$$

$$d_{ij^*} = arg\ max_{k\in\{1,\ldots,J\}}(E(w_{ik}) + z_i\beta_k + u_{ik}) \tag{3.6}$$

Let I_i be a polychotomous variable with values 1 to J and $I_i = j$ if major j is chosen. Following Lee (1983) and Dahl (2002), $I_i = j$ if and only if

$$E(w_{ij}) + z_i\beta_j + u_{ij} \geq max_{k\in\{\{1,\ldots,J\}-j\}} y^*_{ik}$$

where $y^*_{ik} = E(w_{ik}) + z_i\beta_k + u_{ik}$. Then,

$$z_i\beta_j + E(w_{ij}) \geq \xi_{ij} \equiv max_{k\in\{\{1,\ldots,J\}-j\}} y^*_{ik} - u_{ij} \tag{3.7}$$

We prefer to follow a maximum likelihood estimation framework given that marriage is a binary response variable and major selection is a multinomial choice. Adopting normal and logistic distributions are standard in the literature, while, our model additionally takes selection-on-unobservables into account and offer an estimation methodology accordingly.

Constructing the likelihood function in this environment requires to derive $Pr(m_i = 0, I_i = j|x_i, z_i, w_i)$ and $Pr(m_i = 1, I_i = j|x_i, z_i, w_i)$. We assume that the marginal distribution of ε_i denoted by $F_1(\cdot)$ is standard normal $\Phi(\cdot)$. The marginal distribution of ξ_{ij}, $F_2(\cdot)$, is considered under two specifications. The results are either derived by assuming logistic distribution or approximated assuming a joint normal for u_{ik}, $k = 1,\ldots,J$. Then for the joint distribution of ε and ξ denoted by

$F(\varepsilon, \xi)$, by Sklar's result Sklar (1959), there exits a unique copula function $C(\cdot, \cdot)$ s.t. $F(\varepsilon, \xi) = C(F_1(\varepsilon), F_2(\xi))$.[12] Then,

$$\begin{aligned} Pr(m_i = 0, I_i = j | x_i, z_i, w_i) =& Pr(\varepsilon_i \leq -x_i\gamma - \tau_j, \xi_{ij} \leq z_i\beta_j + E(w_{ij})) \\ =& C(F_1(-x_i\gamma - \tau_j), F_2(z_i\beta_j + E(w_{ij}))) \end{aligned}$$

and,

$$\begin{aligned} Pr(m_i = 1, I_i = j | x_i, z_i, w_i) =& Pr(I_i = j | x_i, z_i, w_i) - Pr(m_i = 0, I_i = j | x_i, z_i, w_i) \\ =& Pr(I_i = j | x_i, z_i, w_i) - C(F_1(-x_i\gamma - \tau_j), F_2(z_i\beta_j + E(w_{ij}))) \\ =& F_2(z_i\beta_j) - C(F_1(-x_i\gamma - \tau_j), F_2(z_i\beta_j + E(w_{ij}))) \end{aligned}$$

There are two more steps are needed to derive an estimable equation. The first one is the derivation of $F_2(z_i\beta_j + E(w_{ij}))$. If we assume that u_{ik}, $k = 1, \ldots, J$ are distributed with jointly normal, we can use GHK simulator after Geweke, Hajivassiliou, and Keane, which is found to be the most reliable method for simulating normal rectangle probability among others (e.g., Hajivassiliou et al. (1996)). Following Train (2009),

$$\begin{aligned} F_2(z_i\beta_j + E(w_{ij})) =& Pr(\xi_{ij} < z_i\beta_j + E(w_{ij})) \\ =& Pr(z_i\beta_j + E(w_{ij} + u_{ij} \geq z_i\beta_k + E(w_{ik}) + u_{ik} \; \forall k \neq j) \\ =& Pr(\tilde{V}_{ikj} + \tilde{u}_{ikj} \leq 0 \; \forall k \neq j) \\ =& \int 1(\tilde{V}_{ikj} + \tilde{u}_{ikj} \leq 0 \; \forall k \neq j) \phi(\tilde{u}_{ij}) d\tilde{u}_{ij} \\ =& \int_{\tilde{u}_{ij} \in A_{ij}} \phi(\tilde{u}_{ij}) d\tilde{u}_{ij} \end{aligned}$$

[12]Lee (1983) uses a Gaussian copula.

where $\tilde{V}_{ikj} = z_i(\beta_k - \beta_j) + E(w_{ij}) - E(w_{ik})$ and $\tilde{u}_{ikj} = u_{ik} - u_{ij}$. $A_{ij} \equiv \{\tilde{u}_{ij}\ s.t.\ \tilde{V}_{ikj} + \tilde{u}_{ikj} \leq 0\ \forall k \neq j\}$. This probability is a (J-1) dimensional integral over error differences A_{ij}. This can be approximated by GHK simulator.

Instead of assuming normal distribution and approximating $Pr(\xi_{ij} < z_i\beta_j + E(w_{ij}))$, we can derive $F_2(z_i\beta_j + E(w_{ij}))$ by assuming multinomial logistic (MNL) distribution. MNL is the most common specification in the multinomial choice models. Under MNL distribution, we can write $F_2(\cdot)$ as

$$F_2(z_i\beta_j) = \frac{e^{z_i\beta_j + E(w_{ij})}}{1 + \sum_{j=1}^{J-1} e^{z_i\beta_j + E(w_{ij})}}$$

Second, we need to derive $Pr(m_i = 0, I_i = j|x_i, z_i, w_i)$ and $Pr(m_i = 1, I_i = j|x_i, z_i, w_i)$ using copula and $F_2(z_i\beta_j + E(w_{ij}))$. Using Gaussian copula

$$C(F_1(\varepsilon), F_2(\xi)) = B(\Phi^{-1}(F_1(\varepsilon)), \Phi^{-1}(F_2(\xi))|\rho)$$

where B denotes the bivariate normal distribution. This bivariate distribution assumes that the transformed variables are jointly normal with zero means, unit variances, and correlation coefficient ρ. Then,

$$\begin{aligned}
&C(F_1(-x_i\gamma - \tau_j), F_2(z_i\beta_j + E(w_{ij}))) = B(\Phi^{-1}(F_1(-x_i\gamma - \tau_j)), \Phi^{-1}(F_2(z_i\beta_j + E(w_{ij}))))\\
&= B(\varepsilon_i \leq -x_i\gamma - \tau_j, \xi_{ij} \leq \Phi^{-1}(F_2(z_i\beta_j + E(w_{ij})))|x, z, w)\\
&= \Phi(\varepsilon_i \leq -x_i\gamma - \tau_j|x, z, w)\ \Phi(\xi_{ij} \leq \Phi^{-1}(F_2(z_i\beta_j + E(w_{ij})))|\varepsilon_i \leq -x_i\gamma - \tau_j, x, z, w)\\
&= \Phi(-x_i\gamma - \tau_j)\ \frac{1}{1 - \Phi(x_i\gamma + \tau_j)} \int_{-\infty}^{-(x_i\gamma + \tau_j)} \Phi[(\Phi^{-1}(F_2(z_i\beta_j + E(w_{ij}))) + \rho_j\varepsilon_i)/(1 - \rho_j^2)^{1/2}]\phi(\varepsilon)d\varepsilon\\
&= \int_{-\infty}^{-(x_i\gamma + \tau_j)} \Phi[(\Phi^{-1}(F_2(z_i\beta_j + E(w_{ij}))) + \rho_j\varepsilon_i)/(1 - \rho_j^2)^{1/2}]\phi(\varepsilon)d\varepsilon
\end{aligned}$$

Since we assume that $F_1(\cdot) = \Phi(\cdot)$, $\Phi^{-1}(F_1(\varepsilon)) = \varepsilon$. Then, we can write the likelihood function based on derivations we obtained above such that

$$logL(\beta,\gamma,\tau,\rho) = \sum_{i=1}^{N}\sum_{j=1}^{J} \{1(I_i = j)\ log \int_{-\infty}^{-(x_i\gamma+\tau_j)} \Phi[(\Phi^{-1}(F_2(z_i\beta_j + \tilde{w}_{ij})) + \rho_j\varepsilon_i)/(1-\rho_j^2)^{1/2}]\phi(\varepsilon)d\varepsilon\}^{(m_i=0)}$$

$$\{1(I_i = j)\ log[F_2(z_i\beta_j + \tilde{w}_{ij}) - \int_{-\infty}^{-(x_i\gamma+\tau_j)} \Phi[(\Phi^{-1}(F_2(z_i\beta_j + \tilde{w}_{ij})) + \rho_j\varepsilon_i)/(1-\rho_j^2)^{1/2}]\phi(\varepsilon)d\varepsilon]\}^{(m_i=1)}$$

where $\tilde{w}_{ij}$ is the predicted wage for student i when she graduates from the major j.

Therefore, the model can be estimated by maximum likelihood methods. The parameters $\theta = \{\beta,\gamma,\tau,\rho\}$ estimated by

$$\hat{\theta} = argmax_{\theta} \sum_{i}^{N} logL(\beta,\gamma,\tau,\rho)$$

3.4 Data

We estimate our econometric model using the National Longitudinal Study of Youth 97 cohort (NLYS97). The NLSY97 Cohort is a longitudinal project that follows the lives of a sample of American youth born between 1980 and 1984. There are 8,984 respondents in the sample that were aged 12-17 when they were first interviewed in 1997. This ongoing cohort has been surveyed 15 times until 2013 and is now interviewed biennially.[13] The project follows the participants' life starting from early ages, and therefore we can observe their schooling information, which includes college major choices, after school histories, and marriage related information. Furthermore, this database contains many individual covariates that are used to control observable heterogeneities in our estimation procedure such as gender, age, race, regional characteristics, school performance, earned income, and school characteristics.

We restrict our sample to survey participants who are graduated from higher education institutions. This restriction is compatible with the purpose of our study. Even though there are 8984 survey participants, the number of students that we have full information about the college major

[13] The US Bureau of Labor Statistics https://www.nlsinfo.org

choice is only 1193, which is around 17 percent of the whole sample.[14] Table 3.1 presents the characteristics of male and female students in the data. It is important to note that the age structure of our sample. NLYS97 follows people born between 1980 and 1984, meaning they were 29-33 when the last interview was conducted in 2013.

The female ratio of college graduates is 56% in our sample, which is close to the population ratio (58%).[15] The male participants of the NLYS97 sample is slightly higher than the females, which may account for some differences in the graduate ratios. In terms of marriage outcomes, there are observable differences between female and male graduates. The marriage rate is 5% higher for female graduates and the first marriage for females is almost 10 months earlier. Under the assumption that marriage is a legal contract between a man and a woman[16], we expect to see similar marriage outcomes between men and women. Heterogeneity in the time of the marriage can be one explanation for this difference. In other words, male college graduates have higher tendency to marry later than mid 30's.[17]

Students' performance measures are important in the major choice analysis. Stinebrickner and Stinebrickner (2014) show the differences between the initial beliefs at the time of the college entrance about their final majors and the realized major choices. They find that the main reason of this difference is the misperception about students' own abilities to perform well. Therefore, it is important to control students' performances in first semesters in major choice. From the observed performances in college, there are differences between male and female students. In particular, female students have higher GPAs in the higher education. Similar pattern can also be observed in the first three terms of college education.[18] Performance measures are not only indicators of ability for students to understand their capabilities, but also help them to identify their preferences and talents, which provide directions for their major choices. Therefore, we include these measures in

[14] Note that the participation rate for higher education is lower than the population rate which is around 25% for the 2000's. Because we only consider college graduates, the participation rate difference does not create problem in our setting.

[15] National Center for Education Statistics, Digest of Education Statistics (2012).

[16] Data are also consistent with this definition too.

[17] The fertility concerns of females may lead them to marry earlier Zhang (2014).

[18] Similar controls also can be done with college admissions exam score like SAT and ACT and high school performance measures. We do not include these variables in our analysis because of lack of sufficient data.

Table 3.1: Summary statistics for all college graduates

	Female(56%)		Male(44%)		Total =1193			
Variables	Mean	Sd	Mean	Sd	Mean	SD	Min	Max
Marriage Dummy	0.61	0.49	0.55	0.50	0.59	0.49	0	1
Birth Year	2.65	1.36	2.70	1.43	2.67	1.39	0.08	5
Black	0.21	0.41	0.16	0.36	0.19	0.39	0	1
Hispanic	0.16	0.36	0.14	0.34	0.15	0.36	0	1
White	0.62	0.49	0.69	0.46	0.65	0.48	0	1
Second Degree	0.23	0.42	0.19	0.39	0.21	0.41	0	1
Second College	0.05	0.22	0.06	0.24	0.06	0.23	0	1
Two Year	0.13	0.33	0.13	0.33	0.13	0.33	0	1
Graduate School	0.01	0.09	0.02	0.12	0.01	0.11	0	1
Rural	0.68	0.47	0.69	0.46	0.68	0.47	0	1
North East	0.19	0.39	0.21	0.40	0.20	0.40	0	1
West	0.18	0.38	0.16	0.37	0.17	0.38	0	1
South	0.37	0.48	0.33	0.47	0.35	0.48	0	1
GPA Term 1	2.75	1.11	2.69	1.06	2.72	1.09	0	4
GPA Term 2	2.65	1.15	2.67	0.97	2.66	1.07	0	4
GPA Term 3	2.66	1.17	2.59	1.07	2.63	1.13	0	4
Overall GPA	2.95	0.72	2.83	0.70	2.90	0.72	0	4
Business	0.17	0.38	0.20	0.40	0.18	0.39	0	1
Education	0.10	0.31	0.03	0.18	0.07	0.26	0	1
Computer & Eng.	0.03	0.17	0.15	0.35	0.08	0.27	0	1
Health	0.13	0.34	0.03	0.18	0.09	0.28	0	1
Income	2.08	1.20	2.54	1.51	2.28	1.36	0	13.12

our analysis.

In order to give a picture of the differences between college graduates and the remaining sample, we present summary statistics of full sample in Table A.43, which can be found in the Appendix. There are racial differences in terms of college participation. Whites' college participation rate is higher than Blacks and Hispanics. Also, students from the North have higher fraction of college graduates compared to students from the South and the West. Finally, if we look at marriage outcomes, college graduates marry earlier and more in comparison to the non-college graduate survey participants.

An important factor in major choice is the pecuniary motivation, which is examined extensively for both the college participation decision and college major choice. Its significance is documented in many different contexts. In order to see a general picture for college majors' earning differences,

Table 3.2: Earnings, full time working rates and marriage rates by college major

	Wage			Full time working rates			Marriage Rate		
Majors	Female	Male	Total	Female	Male	Total	Female	Male	Total
Business	10.85	11.18	11.04	0.83	0.91	0.88	0.60	0.54	0.57
Comp & Eng	11.06	11.31	11.25	0.90	0.95	0.94	0.61	0.59	0.60
Education	10.46	10.72	10.54	0.76	0.84	0.79	0.73	0.76	0.74
Health	10.82	11.16	10.88	0.75	0.88	0.77	0.66	0.56	0.65
Others	10.65	11.02	10.83	0.78	0.88	0.83	0.58	0.54	0.56
Total	10.73	11.14	10.95	0.80	0.91	0.86	0.60	0.56	0.58

we additionally present summary statistics from the National Study of College Graduates (NSCG) surveys from 2003, 2010, 2013, which give statistics from larger samples. The National Survey of College Graduates is a longitudinal biennial survey conducted since the 1970s that provides data on the nation's college graduates. These surveys sample individuals who are living in the United States during the survey reference week, have at least a bachelor's degree, and are under the age of 76. Table 3.2 presents the log of the average earnings of college majors constructed from the NSCG data. Computer Science and Engineering graduates have the highest earnings on average. Education major graduates earnings are the lowest among all majors. Women have lower average earnings in all majors but gender differences are slightly different between majors.

Table 3.2 also presents full time working rates and marriage rates from the NSGC data. Working rates are lower in Education and Health major graduates and they are less than 80 percent. Computer science and Engineering, however, are the college majors that have the highest working rates. Full time working rate of college graduates from these majors is more than 94 percent. Women's full time working rates are lower than men's independent of their majors. Marriage rates of survey participants who are younger than 35 are presented in table 3.2. The marriage rates are similar with the NLYS97 data across majors but the rates are slightly higher in the the NLYS97 than the NSCG.

College majors in the NLYS97 are coded according to 2010 College Course Map.[19] We record students' majors according to their transcripts. In the case of a student has more than one college

[19]CCM, http://nces.ed.gov/pubs2012/2012162rev.pdf

transcript, we chase the major which has the highest number credits. Table A.40 presents the major choices of students in the sample. Business and related studies, Liberal Arts and Humanities, Health and related programs, Education, Engineering, and Computer sciences are the most popular majors in our sample. The dispersion of college majors in our sample is similar to the dispersion in the population of college graduates. In the Appendix, Table A.42 presents fractions of college majors among the US college graduates are shown for the years between 2001 and 2014.

In college graduate population, the distribution of majors between male and female students shows divergences. There are many fields dominated by one gender and these patterns have not changed over years. Similar divergences are also seen in our sample. In Education, Engineering, Computer and Information Technologies, Health, and Business related programs the percentage of female and male graduate differences are easily discernible from Table A.40. Business, Engineering and related fields, Computer and Information Technologies are dominated by males, while female students choose Education, Health and related fields and Liberal Art, Humanities majors more.[20] These ratios are not exactly the same but the differences in distribution of majors across gender corresponds to population level differences.[21]

The marital status and the time of the marriage in terms of months are recorded in the NLYS97. Table A.41 presents marital status of the NLYS97 participants who are graduated from college. Female graduates have 5 percent higher marriage rate than males.[22] However, never married and not cohabiting ratio is higher for male graduates. As a summary, we can say that male college graduates have lower marriage and cohabitation rate until their early 30's and female college graduates prefer to marry earlier.

Table A.41 shows the average marriage rate and the month of first marriage of college graduates from different majors. According to Table A.41, graduates from Education major have the highest marriage rate on average, 71 percent of the graduates have experienced at least one marriage before

[20] In addition, we find a similar distribution of college-majors in the NSCG data.

[21] For detailed information about gender differences across majors, see Bronson (2014).

[22] The NLYS97 keeps cohabitation information too. Males cohabitation rates are higher than females, but we don't take cohabitation into account our evaluation because of a lack of legal foundation. Moreover, responsibilities of being married and having a child is not common in cohabitation which changes our evaluation.

33. Computer and Information Sciences major graduates have the lowest rate, only 46 percent of graduates had been married by the time of the survey. Also, some of the majors have considerable differences in marriage rates between females and males. The marriage rate of female graduates from Biological sciences is 48 percent, while it is 67 percent for males. An opposite scenario occurs for Business major graduates; females' marriage rate is 63 percent, males' is 50 percent.

In this section, we present the differences in major choice and marriage outcomes of college graduates. The variation in marriage outcomes combined with the variation in the major choices motivate us to examine the relationship of these decisions. However, there are many factors that can affect in both the marriage outcome and college-major choice. In the next section, we will control other characteristics of participants to reach reliable estimates.

3.5 **Results**

3.5.1 **Major choice and marriage realizations under exogeneity**

Table 3.3 presents estimation results for college major choice under the multinomial logistic specification. We focus on 4 majors; Education, Business, Health, and Computer&Engineering because they are the most popular choices in our sample and also show divergences in terms of male-female proportions. In order to control pecuniary motivations, we use counterfactual wage estimates that are predicted using sample data and the additional exogenous variations.[23] We normalize the coefficient of expected earning to 1, therefore we can interpret result in terms of earnings. The gender has the most significant and quantitively large impact on major choices. Female students are significantly less likely to choose Computer Science and Engineering majors, while they are significantly more likely to choose Education and Health majors. Within the first three terms' GPAs, performance measures have only significant effect on Computer Science and Engineering major choice. Students who have higher GPAs prefer Computer Science and Engineering majors more.

[23] The estimation result for earnings is presented by table A.39 in the Appendix

Table 3.3: Multinomial logistic regression results for major choice

	Education		Business		Health		Computer & Eng	
	Est.	Std. err	Est.	Std. err	Est.	Std. err	Est.	Std. err
Female	1.19	0.28	-0.08	0.16	1.44	0.28	-1.52	0.27
GPA term 1	-0.13	0.12	-0.02	0.08	0.04	0.12	0.03	0.12
GPA term 2	0.16	0.14	-0.01	0.09	-0.14	0.12	0.29	0.15
GPA term 3	0.18	0.13	-0.07	0.08	0.09	0.11	-0.09	0.12
Rural	-0.51	0.25	-0.14	0.17	-0.40	0.23	0.28	0.26
West	-0.40	0.36	-0.99	0.29	-0.56	0.37	-0.76	0.38
South	-0.14	0.30	0.00	0.20	-0.07	0.28	0.02	0.29
North East	-0.70	0.37	-0.32	0.23	-0.47	0.32	-0.25	0.32
Black	-0.29	0.38	-0.10	0.22	0.11	0.30	-0.19	0.34
Hispanic	0.05	0.36	0.12	0.24	-0.41	0.39	0.07	0.34
Birth month	0.18	0.09	-0.01	0.06	0.01	0.08	0.02	0.08
Constant	-3.11	0.60	-0.67	0.36	-2.70	0.55	-2.73	0.55

This table presents multinomial logistic regression results of students' major choices. The coefficient of the expected wage is normalized to 1, therefore coefficients can be interpreted as earnings.

The results show that regional differences and city characteristics have significant impacts on college major choice. Students from the rural areas are less likely to choose Education and Health majors. It is also worth noting that, students from the West are significantly less likely to choose Business and Computer & Engineering majors.

If we assume that there are no unobservable factors that affect college major choice and marriage outcome simultaneously or independence of any unobservable factors, we can estimate the marriage outcome equation by using binary response data methods. These results provide a benchmark to see the differences of college graduates with various majors. Table 3.4 presents estimation result under 3 different specifications (LPM, Probit, Logit). There is a significant difference in the marriage realizations between the graduates from Education major and the other majors' graduates. In all specifications, Education major graduates' marriage rates is higher than others.

There are other significant differences in marriage outcomes. Female graduates marry significantly more than male graduates, which can also be interpreted as female graduates are married earlier. Age is another significant determinant of marriage. Marriage realizations is increasing with age.

The regions of student have also significant impacts on marriage outcomes of college graduates.

Table 3.4: Estimation results without selection into account

	LPM		Probit		Logit	
	Estimate	Std. err	Estimate	Std. err	Estimate	Std. err
Female	0.06	0.03	0.17	0.08	0.27	0.13
Birth month	-0.05	0.01	-0.14	0.03	-0.23	0.04
Hispanic	-0.07	0.04	-0.20	0.11	-0.32	0.18
Black	-0.27	0.04	-0.71	0.10	-1.16	0.17
West	-0.03	0.04	-0.09	0.12	-0.14	0.20
South	0.00	0.04	-0.02	0.10	-0.02	0.16
North East	-0.10	0.04	-0.27	0.11	-0.44	0.18
Education	0.13	0.05	0.38	0.16	0.62	0.26
Business	-0.01	0.04	-0.02	0.10	-0.02	0.16
Health	0.04	0.05	0.09	0.14	0.16	0.23
Com& Eng	-0.02	0.05	-0.06	0.14	-0.10	0.23
Constant	0.77	0.04	0.73	0.12	1.17	0.19

This table presents regression results for marriage realizations under 3 specifications; Linear Probability Model, Probit, and Logit

Compared to college graduates from the North Central, college graduates originated from North East postpone their marriages significantly. In terms of race, Blacks marry less compare to Whites, but there is no significant differences between Hispanics and Whites.

Table 3.5 presents gender specific results. The left (right) side of the table shows parameter estimates of marriage realization equation for female (male) college graduates. There is a significant impact of graduating from Education major on the marriage realizations for females. Female college graduates with Education major have significantly higher marriage rates, while there is no similar effect found for male graduates. Regional differences and city level characteristics have stronger impacts on marriage realizations for male graduates. College graduates from the North East postpone their marriages significantly for both genders.

3.5.2 **Testing the effect of marriage expectations on college major choice**

In this section, we apply our proposed econometric procedure that is explained in Section 3 to NLYS97 sample. In the proposed framework, unobservable marriage expectations factor in the college major choice and marriage realizations are allowed to be correlated. In this way, we can

Table 3.5: Regression results of marriage outcome variables separated by gender

	Female			Male		
	LPM	Probit	Logit	LPM	Probit	Logit
Education	0.15	0.46	0.75	0.03	0.08	0.12
	(0.06)	(0.18)	(0.31)	(0.12)	(0.31)	(0.51)
Business	0.04	0.12	0.19	-0.06	-0.17	-0.28
	(0.05)	(0.14)	(0.23)	(0.06)	(0.15)	(0.24)
Health	0.03	0.07	0.12	0.08	0.21	0.37
	(0.06)	(0.16)	(0.26)	(0.12)	(0.33)	(0.55)
Com& Eng	-0.01	-0.01	-0.02	-0.03	-0.08	-0.14
	(0.11)	(0.30)	(0.49)	(0.06)	(0.16)	(0.26)

This table presents separate estimation results for each gender. We only report the estimated coefficients for college majors, however, the same set of regressor that is used in table 3.4 is used.

control the potential effects of marriage expectations on the college major choice and compare the results with the exogenous case. In other words, after controlling potential channels that are expected to have impacts on both college major choices and marriage realizations, if there is no marriage expectations' effects on major choice, the estimation results from proposed econometric framework and the results where selection-on-unobservables are not taken into account are not different from each other. However, if there are underlying marriage expectations that affect college major choices, it is expected to reach different estimates in these models.

Table 3.6 presents the estimation results of the proposed model in Section 3 under 4 different specifications. In all specifications, we adopt probit model for marriage realizations, however, we alternate the major choice specifications with multinomial logistics (MNL) and multinomial probit (MNP). These two models are workhorses of multinomial choice models in the literature. In addition, we present results for two estimation methods that is proposed to control unobserved heterogeneity; two-step and joint maximum likelihood estimations.

In order to control exogenous variations in college major choice, we include students' first three term college GPAs in the college major choice equations, which are not part of the marriage outcome equation. The economics literature documents the significant impact of student performances on their major choices (among others, Stinebrickner and Stinebrickner (2014)), therefore first college GPAs from first three terms provide variation in major choice. Moreover, since these

GPAs are from the first three terms they we can consider these performance measures as independent of the effect of the chosen major.

Table 3.6: Estimation results for marriage realizations with taking unobserved factors in the account

	TS-MNL		Full-MNL		TS-MNP		Full-MNP	
	Est.	Std. err	Est.	Std. err	Est.	Std. err	Est.	Std. err
Female	0.04	0.09	0.03	0.07	0.16	0.09	0.34	0.23
Birth month	-0.10	0.03	-0.09	0.02	-0.13	0.04	-0.08	0.08
Hispanic	-0.13	0.12	-0.08	0.10	-0.20	0.11	-0.14	0.24
Black	-0.63	0.10	-0.46	0.09	-0.70	0.12	-0.60	0.23
West	-0.47	0.11	-0.32	0.10	-0.10	0.14	0.01	0.21
South	-0.04	0.10	-0.04	0.08	-0.02	0.10	0.06	0.23
North East	-0.26	0.10	-0.29	0.09	-0.25	0.12	0.07	0.21
Education	2.05	0.36	2.49	0.14	1.47	0.60	0.77	0.20
Business	0.38	0.88	-0.66	0.22	0.59	0.65	0.24	0.21
Health	1.23	0.51	-0.22	0.58	0.83	0.61	0.58	0.24
Com & Eng	0.91	0.57	1.08	0.66	0.53	0.52	0.20	0.26
Other	1.20	0.21	1.19	0.10	0.69	0.28	0.31	0.24
$\rho_{Education}$	-3.90	3.93	-4.29	1.22	-0.48	0.71	0.04	0.23
$\rho_{Business}$	-0.49	1.64	-2.71	1.00	-0.10	1.17	-0.10	0.20
ρ_{Health}	0.84	0.72	-0.85	0.76	0.04	0.75	-0.06	0.24
$\rho_{Com\&Eng}$	0.74	0.75	0.69	0.88	-0.13	0.55	-0.05	0.27
ρ_{Other}	2.73	3.32	7.14	1.56	0.07	2.33	0.04	0.20

This table presents estimation results of that the proposed estimation framework in Section 2. We adopt 4 specifications; two-step estimation with MNL distribution, joint MLE estimation with MNL distribution two-step estimation with MNP distribution, joint MLE estimation with MNP distribution. For MNP specifications, we take the number of simulations as 100 for the GHK simulator.

The results in the table 3.6 show that chosen majors have significant impacts on marriage outcomes of college graduates. In all 4 specifications, college graduates from Education majors have significantly higher marriage realizations. However, the estimated effects of graduating from other major on marriage realizations are changing with specifications.

Apart from the effects of the observable characteristics of college graduates on marriage outcomes, table 3.6 presents also the vector of correlation coefficient parameters, ρ. In our specification, ρ represents the correlation coefficient of residuals of marriage outcome equation, (ε), and the residuals of polychotomous variable, (ξ), as it defined in the equation 3.7. Even though these parameters enable us the control potential correlation of unobservable marriage expectations, they

Table 3.7: Average treatment effect of college majors

	(1)		(2)		(3)		(4)		(5)	
	Probit		TS-MNL		TS-MNP		Full-MNL		Full-MNP	
	ATE	Std err	ATE	Std err	ATE	Std err	ATE	Std err	ATE	Std err
Education	0.10	0.06	0.50	0.11	0.47	0.16	0.57	0.14	0.27	0.06
Business	-0.01	0.04	0.17	0.27	0.21	0.21	-0.22	0.14	0.09	0.07
Com & Eng	-0.01	0.06	0.30	0.21	0.19	0.18	0.38	0.21	0.07	0.09
Health	0.01	0.06	0.26	0.28	0.29	0.19	0.08	0.28	0.21	0.08

Standard errors are calculated with nonparametric bootstrap simulations with 100 replications

Table 3.8: Average treatment effects (ATE) of college majors

	(1)		(2)		(3)		(4)		(5)	
	Probit		TS-MNL		TS-MNP		Full-MNL		Full-MNP	
	ATE	Std err	ATE	Std err	ATE	Std err	ATE	Std err	ATE	Std err
Education	0.11	0.05	0.25	0.14	0.23	0.18	0.43	0.17	0.15	0.07
Business	-0.04	0.04	-0.16	0.28	-0.09	0.25	-0.56	0.17	-0.08	0.10
Com & Eng	-0.04	0.06	0.00	0.23	-0.11	0.20	0.19	0.23	-0.09	0.09
Health	-0.01	0.06	-0.06	0.29	0.01	0.20	-0.19	0.29	0.07	0.08

Standard errors are calculated with nonparametric bootstrap simulations with 100 replications do not provide a framework to make direct inference from their estimated values.

Comparing the estimates from different specifications cannot be done by estimated coefficients because there are different scaling factors under different specifications. In order to compare these estimates we present average treatment effects (ATE) in tables 3.7 and 3.8. ATEs are comparable because their scaling factors are embedded in the estimation results.

Since college major choice is a multinomial regressor in our specification, there are different ways to define average treatment effects of chosen majors. We define two types of ATE. In table 3.7, ATE of chosen major j is defined as

$$ATE_j = \frac{1}{N}\sum_{i=1}^{N}\Phi(x_i\gamma + \tau_j) - \Phi(x_i\gamma)$$

This definition of ATE is a direct analogy from the binary variable regressors. In addition, we propose another definition for ATE for multiple treatments. Since a student has to choose one major in the multinomial regressors case, comparing the probabilities of choosing with not choosing major j may not represent the real average treatment effect in this model. In order to deal with this problem, we propose an ATE of chosen major j

$$ATE_j = \frac{1}{N}\sum_{i=1}^{N}(\Phi(x_i\gamma + \tau_j) - \frac{1}{J-1}\sum_{k=1,k\neq j}^{J}\Phi(x_i\gamma + \tau_k))$$

In words, we calculate difference in probabilities of chosen major j with the average probabilities of choosing other majors.[24] The first model in tables 3.7 and 3.8 presents ATE estimates where the marriage expectations and major choices are assumed to be exogenous. The models (2), (3), (4), and (5) report ATE estimates under our proposed framework, marriage expectations and major choices are allowed to be correlated. It is clear from the differences between in the magnitude of ATE estimates, the exogeneity assumption for the effects of marriage expectations on college major choice is rejected.

In addition, we estimate ATE of chosen majors for males and females separately to see any

[24] Heckman et al. (2006) define ATE as the difference between selected choice and the counterfactual outcome.

differences in terms of major choice. Tables 3.9 and 3.10 present ATE estimates for only female and male graduates respectively. The differences in ATE estimates between models (1), (2), and (3) are not as big as in the full sample estimates. For both female and male only ATE estimates with the marriage expectations are taken into account are slightly bigger in terms of magnitude, however, they are not statistical significant in most of the cases.

Table 3.9: Average treatment effects (ATE) college majors only for female graduates

	(1)		(2)		(3)	
	Probit		TS-MNL		TS-MNP	
	ATE	Std Err	ATE	Std Err	ATE	Std Err
Education	0.452	0.061	0.482	0.270	0.517	0.221
Business	0.344	0.055	0.355	0.305	0.335	0.222
Computer & Eng	0.297	0.115	0.082	0.352	-0.117	0.256
Health	0.319	0.055	0.333	0.313	0.241	0.216

Standard errors are calculated with nonparametric bootstrap simulations with 100 replications

Table 3.10: Average treatment effects (ATE) college majors only for male graduates

	(1)		(2)		(3)	
	Probit		TS-MNL		TS-MNP	
	ATE	Std Err	ATE	Std Err	ATE	Std Err
Education	0.346	0.120	0.419	0.321	0.474	0.303
Business	0.230	0.057	0.238	0.255	0.263	0.235
Computer & Eng	0.254	0.069	0.290	0.256	0.230	0.250
Health	0.376	0.138	0.292	0.375	0.312	0.339

Standard errors are calculated with nonparametric bootstrap simulations with 100 replications

Finally, it may not be fair to conclude that the relationship between marriage expectations and college major choices is the most powerful factor in the major choice. However, differences in the average treatment effect estimates suggest a significant link. Even after expected earnings and individual characteristics are controlled in the testing procedure, there may be still other unobservable factors (e.g., physical characteristics) accounted for these differences in major choice and marriage realizations. Therefore, applying proposed test to richer data sets provide us more convincing results about the effects of marriage expectations on college major choice.

3.6 Conclusion

In this paper, we examine differences in marriage outcomes of college graduates. We quantify that marriage realizations of college graduates are differentiated according to the majors they are graduated from. To test the effect of marriage expectations on college majors, we develop a procedure to test the effects of marriage expectations on college major choices. Our test rejects the exogeneity of the effects of marriage expectations on college majors choice even after expected earnings and other observable heterogeneities are controlled for. Moreover, we show that the effects of marriage expectations are different between the female and male students. These findings complement to college major choice literature's findings that non-pecuniary factors are the key determinants of college-major choices as in Beffy et al. (2012) and Wiswall and Zafar (2015).

Our findings about the effects of marriage expectations and gender differences in the major bring another dimension into college major choice in college education. From a policy perspective, major choices have direct influence on the skill distribution in the labor market. Therefore, it can be said that any marriage and labor market related conditions that would have an effect on lifestyles could also have significant impacts on college students' career choices, especially for female students. Moreover, by considering the college enrollment rates, college graduated women's labor market participation rates are much lower than those of men's. In line with this fact, policy makers/employers could produce family friendly work environments to attract women to choose specific career paths, increase college graduate women's labor force participation, and balance skill distribution in the labor market.

BIBLIOGRAPHY

Abdulkadiroglu, A. (2013). School choice. *The Handbook of Market Design*.

Abdulkadiroğlu, A., Agarwal, N., and Pathak, P. A. (2015a). The welfare effects of coordinated assignment: Evidence from the nyc hs match. Technical report, National Bureau of Economic Research.

Abdulkadiroğlu, A., Agarwal, N., and Pathak, P. A. (2017). The welfare effects of coordinated assignment: Evidence from the new york city high school match. *American Economic Review*, 107(12):3635–89.

Abdulkadiroğlu, A., Che, Y.-K., and Yasuda, Y. (2015b). Expanding "choice"in school choice. *American Economic Journal: Microeconomics*, 7(1):1–42.

Abdulkadiroglu, A. and Sönmez, T. (2003). School choice: A mechanism design approach. *The American Economic Review*, 93(3):729–747.

Agarwal, N. and Somaini, P. (2016). Demand analysis using strategic reports: An application to a school choice mechanism.

Ajayi, K. F. (2015). School choice and educational mobility: Lessons from secondary school applications in ghana.

Akyol, P. and Krishna, K. (2017). Preferences, selection, and value added: A structural approach. *European Economic Review*, 91:89–117.

Al-Najjar, N. I. and Pomatto, L. (2016). Choice under aggregate uncertainty. *Theory and Decision*, 80(2):187–209.

Alkan, A., Carkoglu, A., Filiztekin, A., and Inceoglu, F. (2008). Trkiye ortaogretim sektorunde katma deger olusumu: niversiteye giris yarisi. Technical report, Sabanci University.

Alptekinoğlu, A. and Semple, J. H. (2016). The exponomial choice model: A new alternative for assortment and price optimization. *Operations Research*, 64(1):79–93.

Arcidiacono, P. (2004). Ability sorting and the returns of to college major. *Journal of Econometrics*, 121:343–375.

Arcidiacono, P. (2005). Affirmative action in higher education: how do admission and financial aid rules affect future earnings. *Econometrica*, 73:1477–1524.

Arslan, H. A. (2018). Preference estimation in centralized college admissions from reported lists. *mimeo*.

Artemov, G., Che, Y.-K., and He, Y. (2017). Strategic “mistakes”: Implications for market design research. Technical report, mimeo.

Attanasio, O. P., Meghir, C., and Santiago, A. (2011). Education choices in mexico: using a structural model and a randomized experiment to evaluate progresa. *The Review of Economic Studies*, 79(1):37–66.

Avery, C. and Levin, J. (2010). Early admissions at selective colleges. *The American Economic Review*, 100(5):2125–2156.

Avery, C. N., Glickman, M. E., Hoxby, C. M., and Metrick, A. (2012). A revealed preference ranking of us colleges and universities. *The Quarterly Journal of Economics*, 128(1):425–467.

Azevedo, E. M. and Budish, E. (2017). Strategy-proofness in the large. Technical report, National Bureau of Economic Research.

Azevedo, E. M. and Leshno, J. D. (2016). A supply and demand framework for two-sided matching markets. *Journal of Political Economy*, 124(5):1235–1268.

Balinski, M. and Sönmez, T. (1999). A tale of two mechanisms: student placement. *Journal of Economic Theory*, 84(1):73–94.

Becker, G. S. (1973). A theory of marriage: Part i. *Journal of Political Economy*, 81(4):13–46.

Beffy, M., Fougere, D., and Maurel, A. (2012). Choosing the field of study in postsecondary education: Do expected earnings matter? *The Review of Economics and Statistics*, 94(1):334–347.

Belzil, C. and Hansen, J. (2002). Unobserved ability and the return to schooling. *Econometrica*, 70(5):2075–2091.

Bettinger, E. P. (2012). Paying to learn: The effect of financial incentives on elementary school test scores. *Review of Economics and Statistics*, 94(3):686–698.

Bicakova, A. and Jurajda, S. (2016). Field-of-study homogamy.

Bičáková, A. and Jurajda, Š. (2017). Gender composition of college graduates by field of study and early fertility. *Review of Economics of the Household*, pages 1–21.

Bodoh-Creed, A. and Hickman, B. R. (2016). College assignment as a large contest.

Bond, T. N., Bulman, G., Li, X., and Smith, J. (2016). Updated expectations and college application portfolios.

Bonhomme, S. and Robin, J.-M. (2010). Generalized non-parametric deconvolution with an application to earnings dynamics. *The Review of Economic Studies*, 77(2):491–533.

Bronson, M. A. (2014). Degrees are forever: Marriage, educational investment, and lifecycle labor decisions of men and women. *Unpublished manuscript*, 2.

Burgess, S., Greaves, E., Vignoles, A., and Wilson, D. (2015). What parents want: School preferences and school choice. *The Economic Journal*, 125(587):1262–1289.

Buser, T., Niederle, M., and Oosterbeek, H. (2014). Gender, competitiveness, and career choices. *The Quarterly Journal of Economics*, 129(3):1409–1447.

Calsamiglia, C., Fu, C., Güell, M., et al. (2014). Structural estimation of a model of school choices: The boston mechanism vs. its alternatives. *Unpublished paper, UAB*.

Calsamiglia, C., Haeringer, G., and Klijn, F. (2010). Constrained school choice: An experimental study. *The American Economic Review*, 100(4):1860–1874.

Caner, A. and Okten, C. (2010). Risk and career choice: Evidence from turkey. *Economics of Education Review*, 29(6):1060–1075.

Caner, A. and Okten, C. (2013). Higher education in turkey: Subsidizing the rich or the poor? *Economics of Education Review*, 35(C):75–92.

Carpena, F., Cole, S., Shapiro, J., and Zia, B. (2017). The abcs of financial education: experimental evidence on attitudes, behavior, and cognitive biases. *Management Science*.

Casalmiglia, C., Fu, C., and Güell, M. (2014). Structural estimation of a model of school choices: The boston mechanism vs. its alternatives. *Documentos de trabajo (FEDEA)*, (21):1–63.

Chiappori, P.-A., Iyigun, M., and Weiss, Y. (2009). Investment in schooling and the marriage market. *American Economic Review*, 99(5):1689–1713.

Dahl, G. B. (2002). Mobility and the return to education: Testing a roy model with multiple markets. *Econometrica*, 70(6):2367–2420.

Demuynck, T. and Seel, C. (2018). Revealed preference with limited consideration. *American Economic Journal: Microeconomics*.

Dillon, E. W. and Smith, J. A. (2017). Determinants of the match between student ability and college quality. *Journal of Labor Economics*, 35(1):45–66.

Drewes, T. and Michael, C. (2006). How do students choose a university? an analysis of applications to universities in ontario, canada. *Research in Higher Education*, 47(7):781–800.

Eckstein, Z. and Wolpin, K. I. (1999). Why youths drop out of high school: The impact of preferences, opportunities, and abilities. *Econometrica*, 67(6):1295–1339.

Fack, G., Grenet, J., and He, Y. (2015). Beyond truth-telling: Preference estimation with centralized school choice.

Fack, G., Grenet, J., and He, Y. (2017). Beyond truth-telling: Preference estimation with centralized school choice and college admissions.

Flabbi, L. and Moro, A. (2012). The effect of job flexibility on female labor market outcomes: Estimates from a search and bargaining model. *Journal of Econometrics*, 168(1):81–95.

Fryer Jr, R. G. (2011). Financial incentives and student achievement: Evidence from randomized trials. *The Quarterly Journal of Economics*, 126(4):1755–1798.

Gale, D. and Shapley, L. S. (1962). College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15.

Ge, S. (2011). Women's college decisions: how much does marriage matter? *Journal of Labor Economics*, 29(4):773–818.

Goodman, S. (2016). Learning from the test: Raising selective college enrollment by providing information. *Review of Economics and Statistics*, 98(4):671–684.

Gourieroux, C., Monfort, A., and Trognon, A. (1984). Pseudo maximum likelihood methods: Theory. *Econometrica: Journal of the Econometric Society*, pages 681–700.

Goussé, M., Jacquemet, N., and Robin, J.-M. (2017). Marriage, labor supply, and home production. *Econometrica*, 85(6):1873–1919.

Haeringer, G. and Klijn, F. (2009). Constrained school choice. *Journal of Economic Theory*, 144(5):1921–1947.

Hafalir, I. E., Hakimov, R., Kübler, D., and Kurino, M. (2016). College admissions with entrance exams: Centralized versus decentralized. Technical report, SFB 649 Discussion Paper.

Hajivassiliou, V., McFadden, D., and Ruud, P. (1996). Simulation of multivariate normal rectangle probabilities and their derivatives theoretical and computational results. *Journal of Econometrics*, 72(1-2):85–134.

Hassidim, A., Romm, A., and Shorrer, R. I. (2016). Strategic behavior in a strategy-proof environment. In *Proceedings of the 2016 ACM Conference on Economics and Computation*, pages 763–764. ACM.

Hastings, J., Kane, T. J., and Staiger, D. (2009). Heterogeneous preferences and the efficacy of public school choice. *NBER Working Paper*, 2145.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica: Journal of the Econometric Society*, pages 1251–1271.

He, Y. (2012). Gaming the boston school choice mechanism in beijing. *Manuscript, Toulouse School of Economics*.

Heckman, J. J., Urzua, S., and Vytlacil, E. (2006). Understanding instrumental variables in models with essential heterogeneity. *The Review of Economics and Statistics*, 88(3):389–432.

Heo, E. J. (2017). Financial aid in college admissions: Need-based versus merit-based. *mimeo*.

Hickman, B. R. (2013). Pre-college human capital investment and affirmative action: a structural policy analysis of us college admissions. *University of Chicago. Unpublished*.

Hickman, B. R. and Bodoh-Creed, A. (2015). Human capital investment and affirmative action: A structural policy analysis of us college admissions. *Typescript, University of Chicago Department of Economics*.

Hitsch, G. J., Hortaçsu, A., and Ariely, D. (2010). Matching and sorting in online dating. *American Economic Review*, 100(1):130–63.

Hoxby, C. and Avery, C. (2013). The missing “one-offs”: The hidden supply of high-achieving, low-income students. *Brookings papers on economic activity*, 2013(1):1–65.

Hoxby, C. M. and Turner, S. (2015). What high-achieving low-income students know about college. *The American Economic Review*, 105(5):514–517.

Hwang, I. M. (2015). *A robust redesign of High School Match*. PhD thesis, The University of Chicago.

Iyigun, M. and Walsh, R. P. (2007). Building the family nest: Premarital investments, marriage markets and spousal allocations. *Review of Economics Studies*, 74(2):507–535.

Jagabathula, S. and Rusmevichientong, P. (2016). A nonparametric joint assortment and price choice model. *Management Science*.

Jalava, N., Joensen, J. S., and Pellas, E. (2015). Grades and rank: Impacts of non-financial incentives on test performance. *Journal of Economic Behavior & Organization*, 115:161–196.

Keane, M. P. and Wolpin, K. I. (1997). The career decisions of young men. *Journal of Political Economy*, 105(3):473–522.

Keane, M. P. and Wolpin, K. I. (2010). The role of labor and marriage markets, preference heterogeneity, and the welfare system in the life cycle decisions of black, hispanic, and white women. *International Economic Review*, 51(3):851–892.

Kimya, M. (2018). Choice, consideration sets, and attribute filters. *American Economic Journal: Microeconomics*.

Kirkebøen, L. J. et al. (2012). Preferences for lifetime earnings, earnings risk and nonpecuniary attributes in choice of higher education. *Discussion Papers*, 725.

Kirkeboen, L. J., Leuven, E., and Mogstad, M. (2016). Field of study, earnings, and self-selection. *The Quarterly Journal of Economics*, 131(3):1057–1111.

Krishna, K., Lychagin, S., and Frisancho, V. (2017). Retaking in high stakes exams: Is less more?

Lee, L.-F. (1983). Generalized econometric models with selectivity. *Econometrica: Journal of the Econometric Society*, pages 507–512.

Li, T., Perrigne, I., and Vuong, Q. (2000). Conditionally independent private information in ocs wildcat auctions. *Journal of Econometrics*, 98(1):129–161.

Li, T. and Vuong, Q. (1998). Nonparametric estimation of the measurement error model using multiple indicators. *Journal of Multivariate Analysis*, 65(2):139–165.

Lien, J. W., Zheng, J., and Zhong, X. (2016). Preference submission timing in school choice matching: testing fairness and efficiency in the laboratory. *Experimental Economics*, 19(1):116–150.

Lien, J. W., Zheng, J., and Zhong, X. (2017). Ex-ante fairness in the boston and serial dictatorship mechanisms under pre-exam and post-exam preference submission. *Games and Economic Behavior*, 101:98–120.

Long, B. T. (2004). How have college decisions changed over time? an application of the conditional logistic choice model. *Journal of Econometrics*, 121(1):271–296.

Manski, C. F. (2004). Measuring expectations. *Econometrica*, 72(5):1329–1376.

Manzini, P. and Mariotti, M. (2012). Categorize then choose: Boundedly rational choice and welfare. *Journal of the European Economic Association*, 10(5):1141–1165.

Menzel, K. (2015a). Inference for games with many players. *The Review of Economic Studies*, 83(1):306–337.

Menzel, K. (2015b). Large matching markets as two-sided demand systems. *Econometrica*, 83(3):897–941.

Metcalfe, R., Burgess, S., Proud, S., et al. (2011). *Student effort and educational attainment: Using the England football team to identify the education production function*. Centre for Market and Public Organization Bristol.

Milgrom, P. R. and Weber, R. J. (1985). Distributional strategies for games with incomplete information. *Mathematics of operations research*, 10(4):619–632.

Myong, S. (2016). Need-based aid from selective universities and the achievement gap between rich and poor.

Niederle, M. and Vesterlund, L. (2007). Do women shy away from competition? do men compete too much? *The Quarterly Journal of Economics*, 122(3):1067–1101.

of Labor, S. B. (2012). National longitudinal study of youth. *Dataset*.

Olszewski, W. and Siegel, R. (2016). Large contests. *Econometrica*, 84(2):835–854.

Papke, L. E. and Wooldridge, J. M. (1996). Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Econometrics*, 11:619–632.

Pathak, P. A. and Sönmez, T. (2008). Leveling the playing field: Sincere and sophisticated players in the boston mechanism. *American Economic Review*, 98(4):1636–52.

Rivers, D. and Vuong, Q. H. (1988). Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of Econometrics*, 39(3):347–366.

Saygin, P. O. (2013). Gender differences in college applications: Evidence from the centralized system in turkey.

Sklar, A. (1959). Distribution functions in n dimensions and their margins. *Publications of the Institute of Statistics at the University of Paris*, 8:229–31.

Snyder, T. D. and Dillow, S. A. (2013). Digest of education statistics, 2012. nces 2014-015. *National Center for Education Statistics*.

Sönmez, T. and Ünver, M. U. (2011). Matching, allocation, and exchange of discrete resources. In *Handbook of Social Economics*. Citeseer.

Stinebrickner, R. and Stinebrickner, T. R. (2014). A major in science? initial beliefs and final outcomes for college major and dropout. *Review of Economics Studies*, 81:426?472.

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.

Willis, R. J. and Rosen, S. (1979). Education and self-selection. *Journal of Political Economy*, 87(5):S7–S36.

Wiswall, M. and Zafar, B. (2014). Determinants of college major choice: Identification using an information experiment. *The Review of Economic Studies*, 82(2):791–824.

Wiswall, M. and Zafar, B. (2015). Determinants of college major choice: Identification using an information experiment. *The Review of Economic Studies*, 82(2):791–824.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

Wooldridge, J. M. (2014). Quasi-maximum likelihood estimation and testing for nonlinear models with endogenous explanatory variables. *Journal of Econometrics*, 182(1):226–234.

Wu, B. and Zhong, X. (2014). Matching mechanisms and matching quality: Evidence from a top university in china. *Games and Economic Behavior*, 84:196–215.

Zafar, B. (2011). How do college students form expectations? *Journal of Labor Economics*, 29(2):301–348.

Zafar, B. (2013). College major choice and the gender gap. *Journal of Human Resources*, 48(3):545–595.

Zhang, H. (2014). Premarital investments and multidimensional matching.

APPENDICES

A.1 Mathematical Appendix

A.1.1 Proof of Lemma 1.2

We want to show first for a given student i with score p_i,

$$E_i[\Omega_{ij}(p_i,\sigma)|t_j^{-1}] = 0 \text{ if } E_i[\Omega_{ik}(p_i,\sigma)|t_k^{-1}] = 0 \text{ and } t_j^{-1} > t_k^{-1} \forall j \in C$$

From Assumption 1.2, we know that subjective assignment probabilities are decreasing with previous year cutoff scores. Therefore if $E_i[\Omega_{ik}(p_i,\sigma)|t_k^{-1}] = 0$ and $t_j^{-1} > t_k^{-1}$ we know that $\int \Omega_{ij}(\tilde{p}_i,\sigma;T)dF(T(t_j^{-1},\xi_i)) \leq \int \Omega_{ik}(\tilde{p}_i,\sigma;T)dF(T(t_k^{-1},\xi_i)$. Moreover, $E_i[\Omega_{ij}(p_i,\sigma)|t_j^{-1}] \geq 0$ because it is probability and cannot be less than 0. Then;

$$\begin{aligned}
&E_i[\Omega_{ik}(p_i,\sigma)|t_k^{-1}] \geq E_i[\Omega_{ij}(p_i,\sigma)|t_j^{-1}] \geq 0 \\
&E_i[\Omega_{ij}(p_i,\sigma)|t_j^{-1}] = 0
\end{aligned}$$

■

A.1.2 Proof of Theorem 1.1

According to theorem 1, L_i^* is ranking of the most preferred colleges within student i's consideration set. Then, we need to show that

$$u_i Pr(L_i^*) - \kappa|L_i^*| \geq u_i Pr(L_i) - \kappa|L_i| \;\; \forall L_i \in \mathscr{R}_i$$

where $\mathscr{R}$ is all possible rank-order lists in the consideration set of student i and $Pr(L_i)$ indicates the probability of assignments of colleges in the list L_i. According to the condition in the theorem 1, the number of reported colleges are strictly less than allowed number limit number of colleges in the list i.e., $|L| < K$. There are 3 ways of deviating from listing most preferred colleges in the L_i within student i's consideration set.

1. Adding a college j into list L_i^*
2. Changing order of list L_i^*
3. Listing a college j in L_i^* which is not part of the most preferred colleges in the accessible region.

In order to show that L_i consists of most preferred colleges within the considerable set of student i, we need to show that all these deviations have to be worse off for student i.

Case 1: If student i add a college j in her L_i in as a last choice. We know that from the condition in the theorem 1.1 students have only $|L_i| < K$ number of colleges that she drives positive utility in her consideration set. Therefore, there are two cases that student i can add a college j in her lists which are either a college that she derives non-positive utility or a college does not belong to her consideration set. Therefore,

$$V_{ij} = Pr(c_j \in L_i)u_{ij} \leq 0$$

where $u_{ij} \leq 0$ if she does not prefer college j or she believes that $Pr(c_j \in L_i) = 0$. Either cases expected utility (V_{ij}) of college j in L_i is equal to zero. According to assumption 1.3 there is a cost of adding college in list, total gain of adding a college j is negative.

Case 2: Assume that student i prefers college j over college j' and they are part of the student i' consideration set, where $u_{ij}, u_{ij} > 0$. If student j list $c_{j'}$ before c_j and report a new list L_i' conditioning on everything else remains the same in the list. Then the expected utility of new list L_i' should be higher than expected utility of L_i^* so that this change makes student i better off.

$$Pr(c_{j'} \in L_i')u_{ij'} + Pr(c_j \in L_i')u_{ij} > Pr(c_{j'} \in L_i^*)u_{ij'} + Pr(c_j \in L_i^*)u_{ij} \tag{A.8}$$

where $Pr(c_{j'} \in L_i')$ indicates the probability of assignment of college j' in listed in the L_i'. According to DA algorithm, an agent is matched with a college in her reported list if she rejected from all colleges that are ranked before it in the lists. We know that $Pr(c_j \in L_i^*) = Pr(t_j < p_i|L_i) > Pr(c_j \in L_i') = (1 - Pr(t_{j'} < p_i|L_i)) \times Pr(t_j < p_i|L_i)$. Also, $Pr(c_{j'} \in L_i') = Pr(t_j' < p_i|L_i) > Pr(c_{j'} \in L_i^*) = (1 - Pr(t_j < p_i|L_i)) \times Pr(t_j' < p_i|L_i)$.

Then, we can rewrite equation A.8 as

$$
\begin{aligned}
&Pr(t_j' < p_i|L_i)u_{ij'} + (1 - Pr(t_{j'} < p_i|L_i)) \times Pr(t_j < p_i|L_i)u_{ij} > \\
&Pr(t_j < p_i|L_i)u_{ij} + (1 - Pr(t_j < p_i|L_i)) \times Pr(t_j' < p_i|L_i)u_{ij'} \\
&-Pr(t_{j'} < p_i|L_i)Pr(t_j < p_i|L_i)u_{ij'} > -Pr(t_{j'} < p_i|L_i)Pr(t_j < p_i|L_i)u_{ij} \\
&u_{ij'} > u_{ij}
\end{aligned}
$$

which contradicts with $u_{ij'} < u_{ij}$. Hence, changing order is not better off for student i.

Case 3: In order to show that case 3, let's assume that student i choose college j in her consideration list which is not part of her most preferred colleges. This implies that u_{ij} is lower than all u_{ik} such that $k \in L_i^*$. Therefore, replacing college j with any college k in list decreases expected utility.

A.2 Additional tables and figures for Chapter 1

Table A.11: Correlations of reported colleges' cutoff scores and students exam scores

	Only exam	Only Hs GPA	Total
First Choice	0.5802	0.4918	0.6199
Median	0.61	0.4867	0.6584
Last Choice	0.601	0.4744	0.6419
Score Range	-0.0248	0.0028	-0.03
	Other correlations in the list		
2	0.5892	0.4817	0.6163
3	0.5891	0.4799	0.6144
4	0.5896	0.4791	0.6138
5	0.5879	0.478	0.6076
6	0.5849	0.4745	0.6034
7	0.5852	0.4745	0.6
8	0.5849	0.4751	0.5997
9	0.5844	0.4738	0.5952
10	0.5833	0.4755	0.5946
11	0.5822	0.4797	0.5932
12	0.5814	0.4779	0.5906
13	0.5806	0.4783	0.5872
14	0.5786	0.4771	0.5861
15	0.5779	0.4751	0.5815
16	0.5777	0.4804	0.5818
17	0.575	0.4798	0.5815
18	0.5727	0.4752	0.5784
19	0.5691	0.4762	0.5744
20	0.5678	0.4759	0.5719
21	0.5654	0.4738	0.572
22	0.5629	0.4697	0.5668
23	0.5563	0.4707	0.5684
24	0.5555	0.4728	0.5705

Notes: Correlation of students' 3 different scores and reported colleges' cutoff scores.

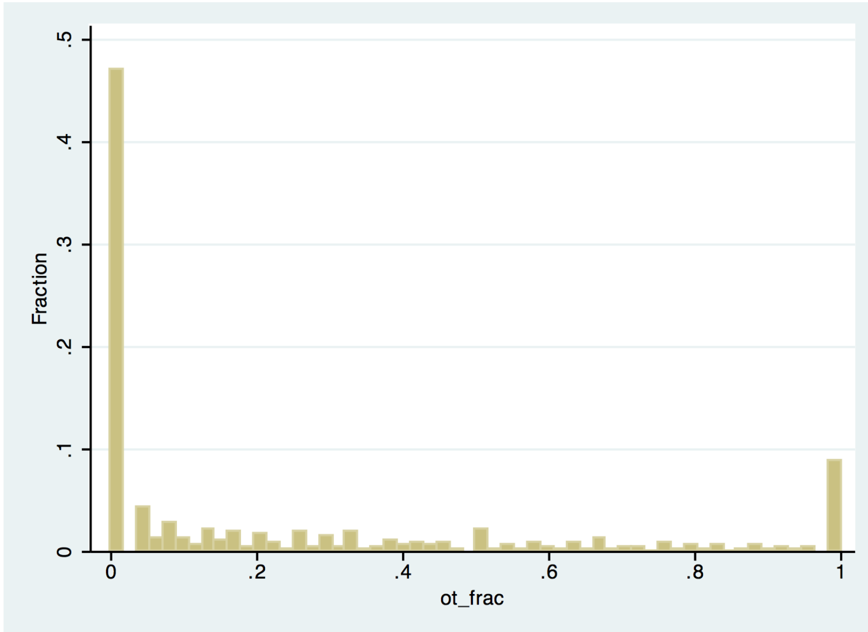


Figure A.1: The fractions of out of track choices in the ROLs

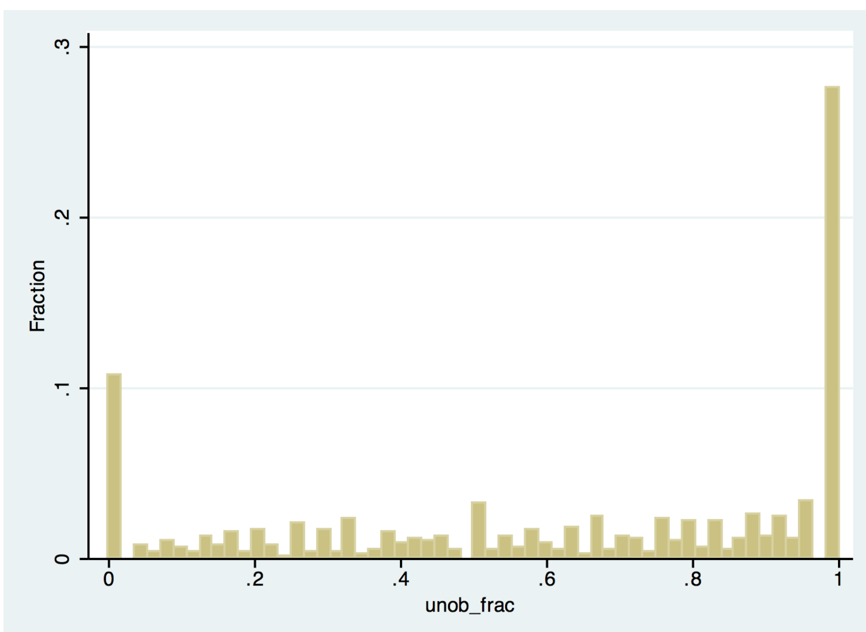


Figure A.2: The fractions of unobtainable choices in the ROLs

Table A.12: Estimation Results: College Dummies Continue

	(Θ_T)		(Θ_S)		(Θ_P)	
	$(Truth - Telling)$		$(Stability)$		$(Cons - Sets)$	
Variables	Coef	St Err	Coef	St Err	Coef	St Err
Female*Private	-0.11	0.09	-0.12	0.11	-0.08	0.10
Female*Distance	0.00	0.01	-0.02	0.01	0.00	0.01
Female*Night	0.13	0.08	0.36	0.10	0.22	0.08
Income*Private	0.31	0.04	0.27	0.05	0.29	0.04
Income*Distance	-0.04	0.00	-0.04	0.00	-0.03	0.00
Income*Night Education	0.08	0.02	0.07	0.03	0.08	0.02
City Size*Private	0.04	0.13	-1.34	0.38	-1.42	0.36
City Size*Distance	-0.07	0.01	-0.07	0.01	-0.08	0.01
City Size*Night Education	-0.19	0.04	-0.23	0.05	-0.22	0.04
Parent Education*Private	-0.33	0.06	-0.44	0.08	-0.38	0.07
Parent Education*Distance	-0.11	0.00	-0.11	0.00	-0.11	0.00
Parent Education*Night E.	0.01	0.04	-0.09	0.06	-0.03	0.05

Notes: This table reports the remaining estimates of college and students' characteristics from table 1.3

Table A.13: Estimation Results: College Dummies

	(Θ_T)		(Θ_S)		(Θ_P)	
	$(Truth-Telling)$		$(Stability)$		$(Cons-Sets)$	
Variables	Coef	St Err	Coef	St Err	Coef	St Err
Adnan Menderes	-0.98	0.12	-0.71	0.16	-0.53	0.12
Afyon Kocatepe	-0.33	0.09	-0.73	0.11	-0.68	0.09
Akdeniz	0.19	0.08	-0.12	0.13	-0.02	0.10
Anadolu	0.59	0.07	0.42	0.09	0.50	0.07
Ankara	0.16	0.08	-0.08	0.12	-0.01	0.10
Ataturk	0.26	0.09	-0.35	0.11	-0.15	0.09
Balikesir	0.27	0.07	0.21	0.09	0.02	0.07
Bogazici	0.34	0.09	1.70	0.12	1.56	0.10
Celal Bayar	-0.25	0.10	-0.63	0.13	-0.53	0.11
Cumhuriyet	-0.05	0.09	-0.64	0.11	-0.55	0.09
On sekiz Mart	0.38	0.07	0.37	0.09	0.26	0.07
Cukurova	-0.17	0.09	-0.47	0.11	-0.47	0.09
Dicle	-0.63	0.11	-1.18	0.14	-1.03	0.12
Dokuz Eylul	0.44	0.08	0.63	0.10	0.63	0.08
Dumlupinar	-0.35	0.09	-0.76	0.12	-0.49	0.09
Ege	0.47	0.08	0.56	0.10	0.52	0.09
Erciyes	-0.10	0.07	-0.42	0.10	-0.32	0.08
Firat	-0.44	0.12	-1.16	0.16	-0.92	0.12
Galatasaray	-0.78	0.14	0.81	0.23	0.59	0.17
Gazi	0.07	0.07	-0.26	0.09	-0.19	0.08
Gaziantep	-0.68	0.13	-1.29	0.16	-1.15	0.14
Gaziosmanpasa	-1.03	0.15	-1.70	0.19	-1.36	0.15
Hacettepe	0.32	0.07	0.25	0.10	0.33	0.08
Harran	-1.43	0.16	-1.74	0.25	-1.31	0.17
Inonu	-0.40	0.10	-0.98	0.12	-0.89	0.10
Istanbul	0.55	0.08	0.64	0.10	0.61	0.08
ITU	0.62	0.09	1.14	0.11	0.96	0.09
Izmir Yuksek Teknoloji	-0.48	0.19	-0.32	0.24	-0.30	0.19
Kafkas	-0.39	0.17	-1.39	0.23	-0.95	0.18
Sutcu Imam	-0.93	0.16	-1.51	0.20	-1.46	0.16
Karadeniz Teknik	0.08	0.09	-0.10	0.10	-0.07	0.09
Kirikkale	-0.44	0.10	-1.16	0.15	-0.87	0.11
Kocaeli	0.35	0.07	0.21	0.09	0.26	0.07
Marmara	0.61	0.08	0.82	0.10	0.78	0.08
Mersin	-0.97	0.09	-0.95	0.12	-0.93	0.10
Mimar Sinan	0.27	0.14	0.34	0.19	0.35	0.15
Mugla	0.04	0.08	0.07	0.10	-0.04	0.08
Mustafa Kemal	-0.79	0.16	-1.60	0.18	-1.38	0.16
Nigde	-0.69	0.11	-0.94	0.14	-0.92	0.11
On Dokuz Mayis	0.26	0.08	-0.19	0.09	-0.08	0.08
ODTU	-0.13	0.09	-0.21	0.12	-0.11	0.10
Osmangazi	0.12	0.08	-0.09	0.10	-0.07	0.09
Pamukkale	-0.05	0.08	-0.20	0.11	-0.19	0.09
Sakarya	-0.05	0.08	-0.32	0.10	-0.15	0.08
Selcuk	0.51	0.06	0.19	0.08	0.22	0.07
Suleyman Demirel	-0.20	0.08	-0.31	0.11	-0.31	0.09

Table A.14: Estimation Results: College Dummies Continue

	(Θ_T)		(Θ_S)		(Θ_P)	
	$(Truth-Telling)$		$(Stability)$		$(Cons-Sets)$	
Variables	Coef	St Err	Coef	St Err	Coef	St Err
Trakya	-0.01	0.08	0.03	0.11	-0.13	0.09
Uludag	0.71	0.07	0.64	0.08	0.62	0.07
Yildiz	0.50	0.09	1.05	0.11	1.04	0.09
Yuzuncu Yil	-0.48	0.17	-0.96	0.21	-0.84	0.18
Karaelmas	-0.35	0.11	-1.07	0.15	-0.90	0.11
Gulhane	0.21	0.47	2.92	0.69	0.58	0.54
Atilim	-0.93	0.57	2.04	1.10	0.51	0.95
Bahcesehir	0.45	0.55	3.83	1.08	2.12	0.93
Baskent	-0.31	0.55	3.02	1.08	1.52	0.93
Beykent	-0.65	0.57	2.28	1.10	0.89	0.94
Bilkent	0.57	0.54	4.63	1.07	2.69	0.93
Cag	-1.59	0.65	-1.94	1.08	-2.68	0.77
Cankaya	-0.27	0.57	3.02	1.10	1.39	0.94
Dogus	-0.64	0.56	2.40	1.09	0.82	0.94
Fatih	-0.92	0.56	2.99	1.10	1.26	0.94
Halic	-0.51	0.57	2.67	1.09	0.93	0.95
Isik	-0.06	0.56	3.26	1.09	1.33	0.94
Bilgi	0.85	0.55	4.51	1.08	2.72	0.93
Kultur	0.04	0.56	3.33	1.08	1.52	0.94
Kadir Has	-0.06	0.56	3.37	1.08	1.56	0.94
Koc	0.56	0.55	5.29	1.07	3.14	0.93
Maltepe	-0.39	0.56	2.90	1.09	1.04	0.94
Sabanci	0.22	0.55	5.53	1.08	3.47	0.93
Yeditepe	0.32	0.55	3.70	1.07	2.00	0.93
Izmir Ekonomi	-0.35	0.56	3.97	1.08	2.01	0.93
Ticaret	0.39	0.55	3.91	1.08	2.13	0.93
Ufuk	-1.55	0.61	1.69	1.13	-0.13	0.96
Yasar	-1.66	0.72	2.26	1.18	0.00	1.05
Okan	-0.23	0.58	2.44	1.10	0.98	0.96
TOBB	0.86	0.53	5.39	1.07	3.23	0.93
Cyprus	2.60	0.64	4.10	0.99	2.44	0.86
Other Abroad	-0.04	0.77	0.85	1.08	0.26	0.92

Table A.15: Estimation Results with students reports complete lists

	(Θ_T)		(Θ_S)		(Θ_P)		(Θ_R)	
	$(Truth-Telling)$		$(Stability)$		$(Cons-Sets)$		$(Only-Report)$	
Variables	Coef	St Err	Coef	St Err	Coef	St Err	Coef	St Err
Same City	0.60	0.04	0.75	0.05	0.67	0.04	-0.05	0.05
Partial Sch.	-0.63	0.10	-0.20	0.14	-0.12	0.10	-1.33	0.13
Full Sch.	-0.77	0.07	0.95	0.10	0.51	0.08	-0.21	0.18
Distance Education	0.68	0.10	-0.95	0.15	-0.66	0.11	2.79	0.28
Night Education	-0.94	0.24	-1.10	0.34	-0.95	0.24	0.35	0.27
Private College	-2.80	0.72	-1.60	1.25	-0.06	1.02	-5.33	2.14
Female*Private	-0.10	0.09	-0.13	0.10	-0.10	0.09	-0.27	0.12
Female*Distance	0.00	0.01	-0.01	0.01	-0.01	0.01	-0.05	0.01
Female*Night	0.18	0.07	0.39	0.09	0.23	0.07	0.03	0.08
Income*Private	0.30	0.03	0.28	0.04	0.28	0.04	0.03	0.03
Income*Distance	-0.04	0.00	-0.04	0.00	-0.03	0.00	-0.06	0.00
Income*Night	0.09	0.02	0.08	0.03	0.09	0.02	-0.02	0.02
City Size*Private	0.11	0.11	-0.94	0.37	-1.00	0.33	1.23	0.63
City Size*Distance	-0.07	0.01	-0.07	0.01	-0.08	0.01	-0.05	0.00
City Size*Night	-0.20	0.03	-0.24	0.04	-0.22	0.03	0.01	0.04
Parent Educ*Private	-0.39	0.06	-0.49	0.08	-0.43	0.07	-0.04	0.07
Parent Educ*Distance	-0.12	0.00	-0.11	0.00	-0.12	0.00	-0.13	0.00
Parent Educ*Night	0.06	0.04	-0.03	0.05	0.01	0.04	0.03	0.04

Notes: This table reports the estimates of college and students' characteristics when the students reported 24 college-majors in their ROLs also taken accounted in the estimation under 4 different methods.

Table A.16: Estimation Results with students reports complete lists

	(Θ_T)		(Θ_S)		(Θ_P)		(Θ_R)	
	$(Truth - Telling)$		$(Stability)$		$(Cons - Sets)$		$(Only - Report)$	
Variables	Coef	St Err	Coef	St Err	Coef	St Err	Coef	St Err
Adnan Menderes	-0.80	0.09	-0.58	0.12	-0.44	0.09	0.03	0.16
Afyon Kocatepe	-0.22	0.07	-0.61	0.08	-0.58	0.07	0.26	0.12
Akdeniz	0.06	0.07	-0.25	0.11	-0.17	0.08	-0.26	0.12
Anadolu	0.44	0.06	0.28	0.08	0.36	0.06	-0.46	0.09
Ankara	-0.07	0.07	-0.29	0.10	-0.29	0.08	-1.17	0.10
Ataturk	0.32	0.07	-0.25	0.09	-0.09	0.08	0.32	0.10
Balikesir	0.26	0.06	0.14	0.07	0.00	0.06	-0.42	0.09
Bogazici	0.06	0.08	1.58	0.11	1.29	0.09	-2.43	0.13
Celal Bayar	-0.16	0.08	-0.50	0.10	-0.42	0.08	-0.17	0.12
Cumhuriyet	0.04	0.07	-0.60	0.09	-0.49	0.07	0.06	0.12
On sekiz Mart	0.40	0.05	0.38	0.07	0.28	0.06	-0.08	0.10
Cukurova	-0.23	0.07	-0.51	0.09	-0.54	0.08	-0.52	0.12
Dicle	-0.56	0.10	-1.07	0.11	-0.95	0.10	0.19	0.15
Dokuz Eylul	0.25	0.06	0.56	0.08	0.47	0.07	-1.07	0.09
Dumlupinar	-0.26	0.07	-0.58	0.09	-0.43	0.07	0.28	0.14
Ege	0.31	0.07	0.49	0.09	0.36	0.07	-0.91	0.10
Erciyes	-0.16	0.06	-0.47	0.08	-0.39	0.07	-0.11	0.10
Firat	-0.38	0.10	-1.09	0.13	-0.87	0.10	0.41	0.18
Galatasaray	-1.02	0.13	0.74	0.21	0.43	0.15	-2.53	0.25
Gazi	-0.06	0.06	-0.29	0.08	-0.31	0.06	-0.87	0.09
Gaziantep	-0.75	0.11	-1.30	0.14	-1.21	0.11	0.07	0.15
Gaziosmanpasa	-0.89	0.12	-1.44	0.15	-1.26	0.12	0.09	0.24
Hacettepe	0.08	0.06	0.10	0.09	0.09	0.07	-1.28	0.10
Harran	-1.32	0.12	-1.57	0.20	-1.22	0.13	1.65	0.32
Inonu	-0.34	0.08	-0.88	0.10	-0.84	0.09	0.29	0.13
Istanbul	0.31	0.07	0.48	0.09	0.37	0.07	-1.29	0.09
ITU	0.39	0.08	0.99	0.10	0.75	0.08	-1.61	0.11
Izmir yuksek Teknoloji	-0.52	0.15	-0.15	0.19	-0.33	0.16	-0.76	0.25
Kafkas	-0.06	0.13	-1.05	0.18	-0.65	0.14	0.54	0.25
Sutcu Imam	-0.95	0.13	-1.38	0.16	-1.50	0.13	0.29	0.22
Karadeniz Teknik	0.11	0.08	0.01	0.08	0.00	0.07	0.11	0.10
Kirikkale	-0.50	0.08	-1.14	0.12	-0.96	0.09	0.06	0.16
Kocaeli	0.27	0.06	0.17	0.07	0.16	0.06	-0.29	0.09
Marmara	0.37	0.07	0.70	0.09	0.56	0.07	-1.25	0.09
Mersin	-0.93	0.08	-0.93	0.10	-0.89	0.08	0.09	0.12
Mimar Sinan	0.06	0.12	0.12	0.18	0.09	0.13	-1.52	0.24
Mugla	0.10	0.07	0.13	0.09	0.03	0.07	0.12	0.11
Mustafa Kemal	-0.72	0.13	-1.42	0.15	-1.30	0.13	0.16	0.22
Nigde	-0.63	0.09	-0.90	0.12	-0.87	0.09	0.07	0.15
On Dokuz Mayis	0.25	0.06	-0.18	0.08	-0.10	0.07	-0.19	0.09
ODTU	-0.36	0.07	-0.37	0.11	-0.31	0.09	-1.89	0.12
Osmangazi	0.03	0.07	-0.17	0.09	-0.18	0.07	-0.40	0.11
Pamukkale	0.04	0.06	-0.11	0.09	-0.09	0.07	-0.17	0.11
Sakarya	-0.06	0.06	-0.20	0.08	-0.16	0.06	0.01	0.11
Selcuk	0.39	0.05	0.09	0.07	0.10	0.06	-0.26	0.08
Suleyman Demirel	-0.18	0.06	-0.31	0.08	-0.30	0.07	0.05	0.11

Notes: This table reports the estimates of college and students' characteristics when the students reported 24 college-majors in their ROLs also taken accounted in the estimation under 4 different methods.

Table A.17: Estimation Results with students reports complete lists

	(Θ_T)		(Θ_S)		(Θ_P)		(Θ_R)	
	$(Truth-Telling)$		$(Stability)$		$(Cons-Sets)$		$(Only-Report)$	
Variables	Coef	St Err	Coef	St Err	Coef	St Err	Coef	St Err
Trakya	-0.05	0.07	0.01	0.09	-0.13	0.07	0.13	0.13
Uludag	0.58	0.06	0.54	0.07	0.48	0.06	-0.42	0.09
Yildiz	0.31	0.08	0.95	0.10	0.87	0.08	-1.40	0.11
Yuzuncu Yil	-0.20	0.13	-0.69	0.16	-0.58	0.14	0.26	0.20
Karaelmas	-0.16	0.08	-0.84	0.11	-0.71	0.08	0.22	0.15
Gulhane	-0.20	0.42	1.68	0.65	-0.05	0.47	-0.03	0.77
Atilim	-1.46	0.51	0.29	0.99	-0.60	0.84	0.44	1.15
Bahcesehir	-0.07	0.50	2.13	0.98	1.07	0.83	0.40	1.12
Baskent	-0.86	0.50	1.26	0.98	0.36	0.83	0.27	1.12
Beykent	-1.15	0.51	0.70	0.99	-0.16	0.84	0.70	1.16
Bilkent	-0.03	0.49	2.78	0.97	1.49	0.83	-0.74	1.12
Cag	-1.61	0.56	-2.11	0.93	-2.55	0.66	3.94	1.36
Cankaya	-0.76	0.51	1.29	0.99	0.35	0.84	0.65	1.14
Dogus	-1.10	0.51	0.78	0.98	-0.18	0.84	0.99	1.13
Fatih	-1.47	0.51	1.26	0.99	0.18	0.84	0.32	1.16
Halic	-1.02	0.51	0.97	0.99	-0.12	0.85	1.24	1.14
Isik	-0.53	0.51	1.56	0.99	0.31	0.84	0.93	1.13
Bilgi	0.28	0.50	2.74	0.97	1.60	0.83	-0.01	1.12
Kultur	-0.48	0.50	1.58	0.98	0.45	0.84	1.07	1.12
Kadir Has	-0.58	0.50	1.67	0.98	0.50	0.84	0.54	1.13
Koc	-0.03	0.50	3.49	0.97	1.95	0.83	-1.52	1.12
Maltepe	-0.87	0.51	1.20	0.98	0.01	0.84	1.14	1.13
Sabanci	-0.35	0.50	3.75	0.97	2.28	0.83	-2.60	1.12
Yeditepe	-0.24	0.50	1.93	0.97	0.86	0.83	0.01	1.12
Izmir Ekonomi	-1.00	0.51	2.03	0.98	0.75	0.83	0.46	1.13
Ticaret	-0.18	0.50	2.08	0.98	0.97	0.83	0.34	1.12
Ufuk	-2.04	0.55	0.00	1.02	-1.19	0.86	0.62	1.19
Yasar	-2.24	0.65	0.34	1.09	-1.15	0.94	0.76	1.27
Okan	-0.73	0.53	0.71	1.00	-0.07	0.85	1.25	1.18
TOBB	0.34	0.48	3.56	0.97	2.04	0.83	-0.32	1.12
Cyprus	1.83	0.58	2.42	0.91	1.40	0.75	4.21	1.08
Other Abroad	-0.79	0.69	-1.02	1.00	-0.83	0.81	36.14	5497

Notes: This table reports the estimates of college and students' characteristics when the students reported 24 college-majors in their ROLs also taken accounted in the estimation under 4 different methods.

Table A.18: Correlation table of cutoff scores across years

	2000	2001	2002	2003	2004
2000	1.00				
2001	0.92	1.00			
2002	0.92	0.95	1.00		
2003	0.80	0.84	0.88	1.00	
2004	0.72	0.76	0.80	0.94	1.00

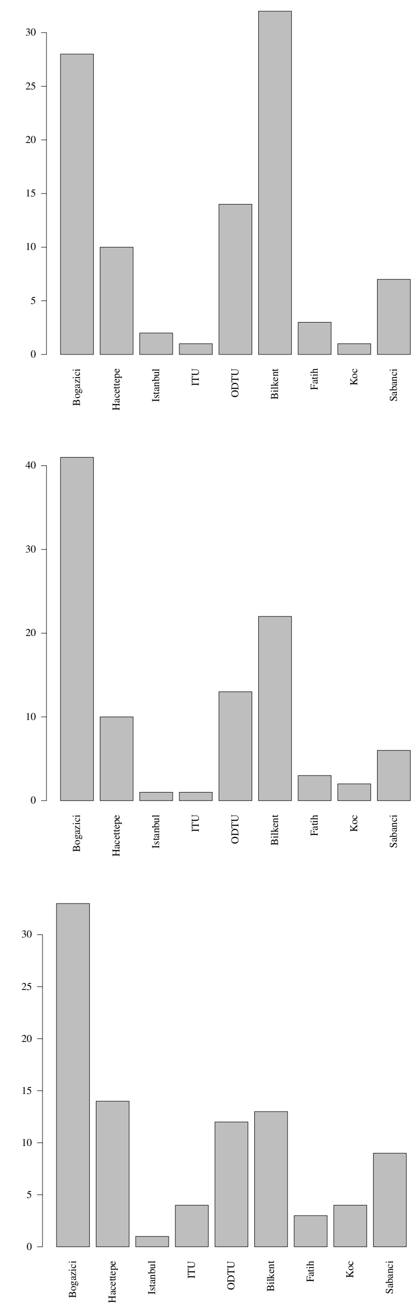


Figure A.3: First 100 ranked students' first choices in ROL by education track

Notes: Each bar-plot shows the first 100 ranked students' first choices according to their education tracks. Quantitative, Verbal, and Equal-weight tracks are ordered from top to bottom.

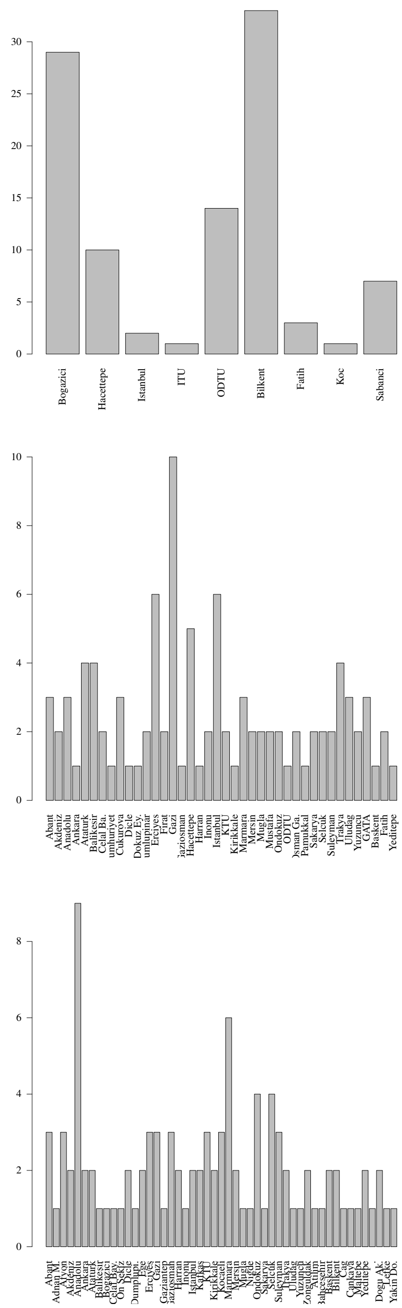


Figure A.4: Students' first choices

Notes: Bar-plot shows the first, median, and last 100 ranked students' first college choices.

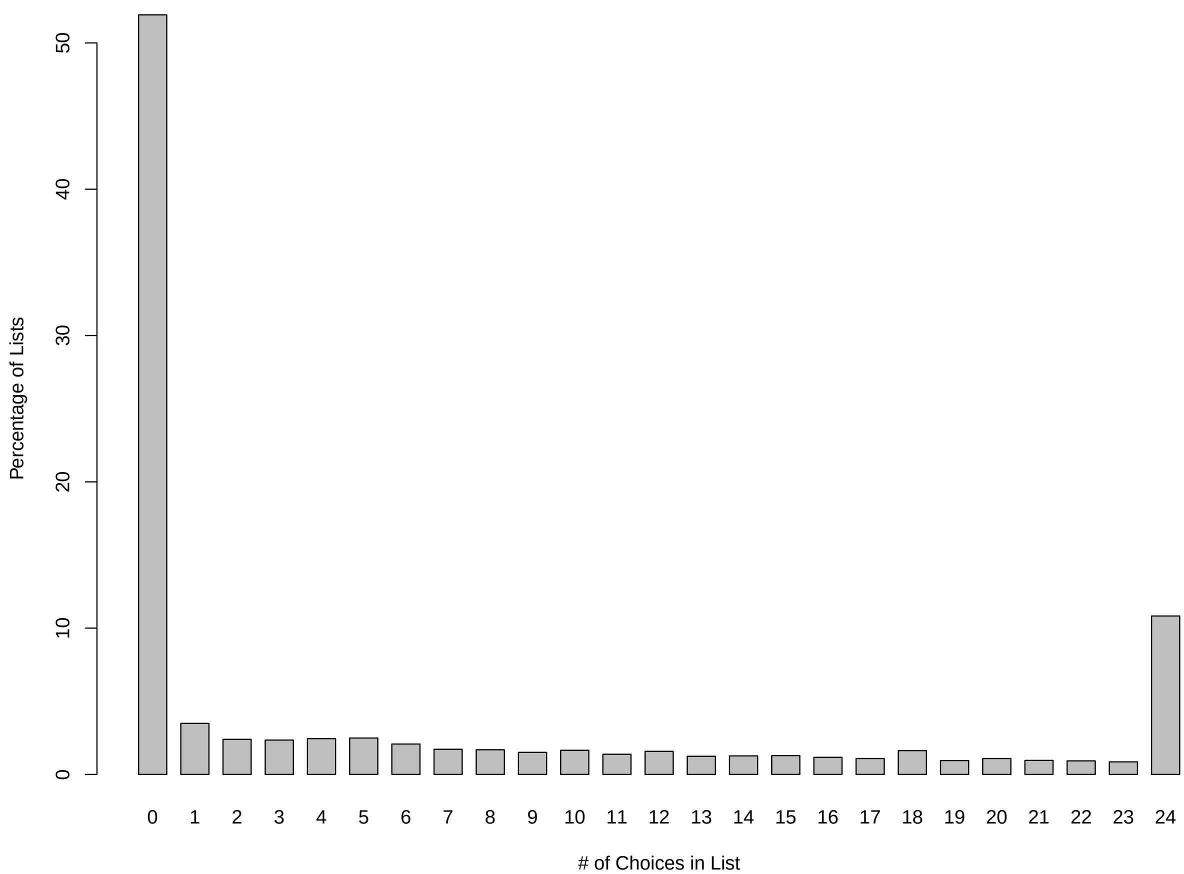


Figure A.5: The percentage of choices in the reported lists

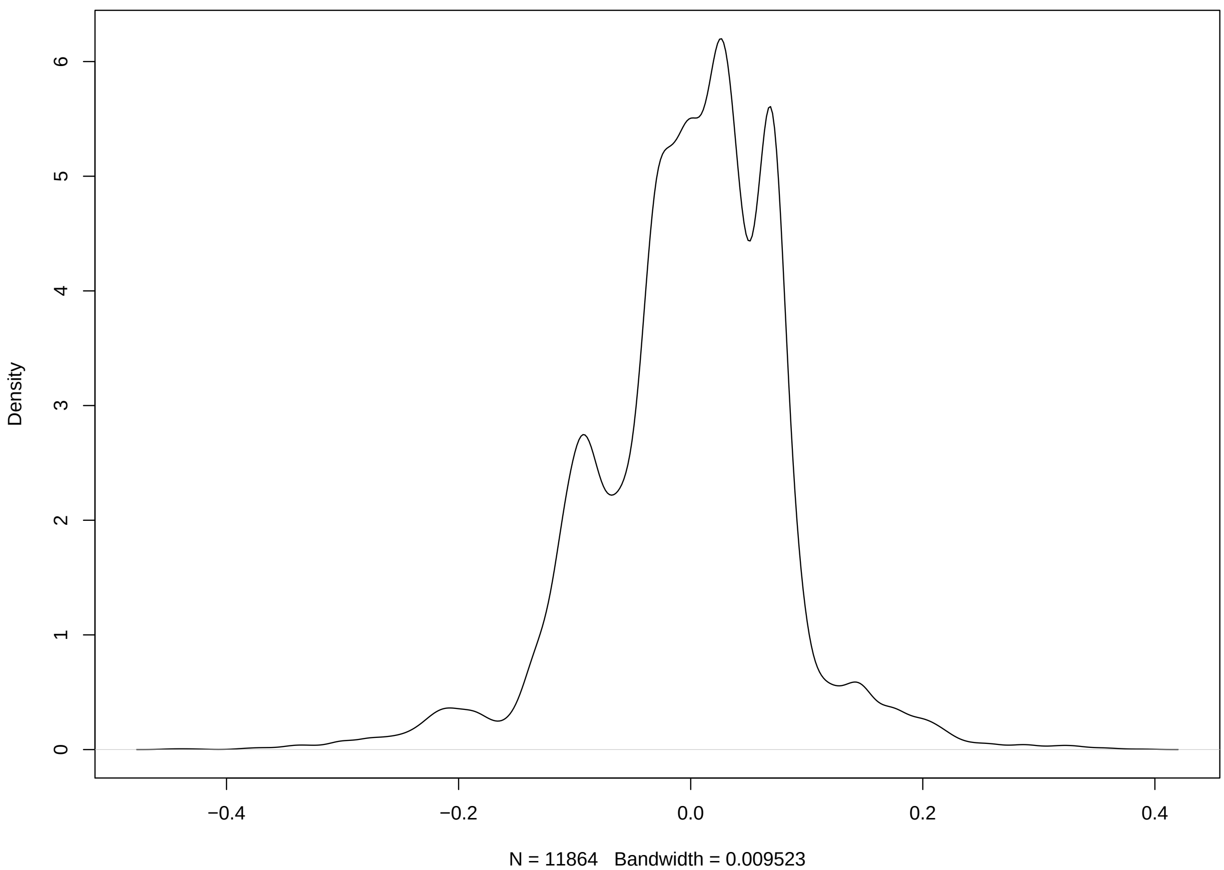


Figure A.6: The density of demeaned cutoff percentiles

A.3 College admission mechanism history in Turkey

A national level college admission started in 1974 after the foundation of the selection and placement center (CSSP). Before the national level admissions, each college accepted students according to application orders, high school field of studies and high school GPAs. With the increase in the number of high school graduates, colleges started to prepare their own entrance exams for sorting the applicants.

The national admission center's mission is to prepare national level test in order to determine applicants' sorting in the college admissions system and by collecting applicants choices match students with the colleges. The center gave two exams for selection and placement. Selection exam chooses student to be placement contest and placement exam is the main determinant for the placement score. From its foundation, there has been several changes in the implementation of the admissions system. In chronological order;

- In 1982, high school GPA's was started to be effective in the college applicants sorting criterion.
- In 1987, college entrance exam scores are differentiated according to field of studies.
- In 1999, the number of college entrance exams reduce to one. Applicants preference lists were started to be submitted after the announcement of college entrance exam scores.
- In 2006, college entrance exam questions started to cover high school education curriculum.
- In 2010, the number of college entrance exams increased to two. The first exam is for selection and qualification exam that selects students to enter placement period and allow successful students to apply two-year degrees. Second exam is the main student sorting determinant for placement.

A.4 Major and location choice in college admissions in Turkey

In this section, I focus major choice in Turkey college admission data in year 2005. Table A.19 presents selected major indicator from rank order lists as fractional responses. In order to present heterogeneity in the major selection, I control many socioeconomic factors of students including parent education levels, income categories, parent occupation categories, and family accumulated wealth. With the exogenous score assumption, in order to control students' strategic effects I include college exam score and scaled high school GPA in the control variables. Estimation results show that there are significant differences in terms of socioeconomic indicators in college major choice apart from exam scores.

Table A.20 presents estimation results from selected location indicator variables obtained from separate OLS regressions. The first model uses fraction of the colleges from three big cities of Turkey as dependent variable. Istanbul, Ankara, Izmir are not only crowded cities, but also they are the big hubs of the commerce and public services of their regions. These cities are includes the prestigious colleges of the country and the number of the colleges are considerably more than the other cities. In addition, since they are the center of commerce and industrial productions, labor markets are bigger in these locations and differentiated from the other smaller cities. Consequently, the cost of living in these big cities are considerably higher than small cities. The results from the first model shows that income and wealth are significant factors in the selection of the colleges in the three big cities of Turkey even the college admissions criteria are controlled.

The second model in Table A.20 shows the estimation results the choice of colleges from the students' cities. In order to eliminate the bigger city effects on choices, I restricted the sample with the students from small cities. Again, my dependent variable is the fraction of colleges from the student's city. Estimation results show that female are inclined more to choose from their own cities and income, wealth are losing their significance in comparison with the bigger cities choices.

Table A.21 presents summary statistics separated by gender. The most interesting differences are the GPA and college exam scores between male and female students. Even though there is substantial differences in favor female students in GPA, college entrance exam score averages

Table A.19: QMLE result for major choice without endogenous college entrance exam score from the survey sample

	Teacher Majors		Medical School		Engineering		Natural Science		Architecture		Math	
	Coef	Std Err	Coef	Std Err	Coef	Std Err	Coef	Std Err	Coef	Std Err	Coef	Std Err
College Exam	0.0001	0.0002	0.0030	0.0002	0.0031	0.0003	-0.0009	0.0002	0.0001	0.0001	0.0008	0.0001
HS GPA	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Mother Education												
Higher Educ.	-0.0905	0.0385	-0.0867	0.0349	0.1002	0.0402	0.0058	0.0219	0.0019	0.0098	0.0050	0.0133
High School	-0.1045	0.0362	-0.0824	0.0326	0.1087	0.0348	0.0239	0.0202	0.0003	0.0084	0.0250	0.0115
Middle School	-0.0366	0.0406	-0.0720	0.0335	0.0614	0.0374	0.0094	0.0214	0.0063	0.0102	0.0330	0.0133
Elementary	-0.0618	0.0349	-0.0614	0.0301	0.0690	0.0309	0.0052	0.0183	-0.0038	0.0065	0.0401	0.0102
Father Education												
Higher Educ.	-0.1100	0.0786	0.0792	0.0450	0.0676	0.0500	-0.0880	0.0577	0.0151	0.0092	-0.0302	0.0174
High School	-0.0916	0.0784	0.0468	0.0447	0.0821	0.0488	-0.0700	0.0578	0.0161	0.0083	-0.0223	0.0170
Middle School	-0.0651	0.0803	0.0438	0.0451	0.0373	0.0502	-0.0675	0.0584	0.0091	0.0088	-0.0211	0.0191
Elementary	-0.0619	0.0780	0.0352	0.0437	0.0302	0.0469	-0.0548	0.0584	0.0032	0.0083	-0.0156	0.0165
Mother Occupation												
Worker	-0.0675	0.0477	0.0160	0.0384	-0.0243	0.0660	0.1048	0.0445	-0.0220	0.0139	-0.0205	0.0235
Official	-0.0215	0.0376	0.0330	0.0263	-0.0343	0.0467	0.0290	0.0213	0.0048	0.0168	0.0150	0.0198
Expert	-0.0519	0.0367	0.0457	0.0261	-0.0337	0.0471	0.0377	0.0232	-0.0123	0.0156	0.0038	0.0188
Small Bus Owner	0.0600	0.0649	0.0420	0.0433	-0.0698	0.0677	-0.0070	0.0229	-0.0067	0.0162	0.0069	0.0305
Big Bus Owner	-0.0154	0.0408	0.0353	0.0496	-0.0694	0.1188	-0.0038	0.0426	-0.0498	0.0221	0.0372	0.0262
Manager High	-0.0110	0.0554	-0.0175	0.0395	-0.0863	0.0934	0.0384	0.0538	0.0373	0.0697	0.0151	0.0207
Manager Middle	-0.0321	0.0722	0.0205	0.0521	-0.0016	0.0961	-0.0154	0.0221	0.0324	0.0365	-0.0263	0.0184
Sport Art	-0.0376	0.0473	-0.0488	0.0369	0.1598	0.0623	-0.0964	0.0326	-0.0463	0.0211	-0.0406	0.0215
Agriculture	-0.0708	0.0786	0.0167	0.0622	-0.1089	0.0774	0.0102	0.0339	0.0032	0.0207	0.0545	0.0601
Retired	-0.0141	0.0383	0.0237	0.0266	-0.0415	0.0479	0.0327	0.0217	-0.0005	0.0177	-0.0096	0.0168
Searching Job	-0.0464	0.1026	-0.0296	0.0368	0.0706	0.1222	0.1584	0.0816	-0.0079	0.0204	0.0006	0.0303
Non wage Income	0.0986	0.0972	0.0912	0.0844	-0.1118	0.0869	-0.0504	0.0224	0.0059	0.0306	-0.0372	0.0188
Housewife	-0.0285	0.0357	0.0346	0.0209	-0.0333	0.0420	0.0153	0.0195	-0.0028	0.0140	0.0023	0.0163
Father Occupation												
Worker	0.0078	0.0294	-0.0047	0.0216	-0.0025	0.0345	-0.0143	0.0206	-0.0075	0.0093	0.0083	0.0170
Official	0.0429	0.0249	0.0116	0.0198	-0.0402	0.0302	-0.0159	0.0167	-0.0128	0.0099	-0.0010	0.0145
Expert	-0.0059	0.0239	0.0074	0.0232	-0.0105	0.0354	0.0155	0.0192	0.0057	0.0133	-0.0031	0.0141
Small Bus Owner	-0.0447	0.0244	0.0258	0.0202	0.0280	0.0326	-0.0073	0.0184	-0.0026	0.0105	-0.0164	0.0138
Big Bus Owner	-0.0320	0.0266	0.0094	0.0251	0.0348	0.0413	0.0045	0.0236	0.0152	0.0164	-0.0237	0.0132
Manager High	0.0007	0.0280	-0.0493	0.0249	-0.0067	0.0417	-0.0002	0.0204	0.0098	0.0191	-0.0114	0.0157
Manager Middle	0.0124	0.0342	-0.0082	0.0302	-0.0434	0.0418	-0.0298	0.0179	0.0085	0.0179	0.0157	0.0200
Sport Art	-0.0412	0.0320	-0.0646	0.0385	-0.0917	0.1232	0.0171	0.0561	0.0167	0.0377	-0.0275	0.0161
Agriculture	0.0592	0.0457	0.0327	0.0320	-0.0052	0.0428	-0.0395	0.0206	-0.0030	0.0126	-0.0401	0.0181
Retired	-0.0016	0.0272	0.0168	0.0221	-0.0034	0.0334	0.0031	0.0194	-0.0043	0.0107	-0.0050	0.0145
Searching Job	0.0246	0.0548	-0.0070	0.0510	0.0138	0.0606	-0.0987	0.0305	-0.0006	0.0129	-0.0072	0.0277
Non wage Income	0.0158	0.0673	-0.0636	0.0348	0.1846	0.1109	-0.0433	0.0353	0.0054	0.0304	0.0468	0.0555
Income Categories												
2	-0.0143	0.0874	-0.0465	0.0702	0.0222	0.0734	-0.0426	0.0357	-0.0080	0.0110	0.0248	0.0358
3	-0.0345	0.0812	-0.0629	0.0620	0.0150	0.0716	0.0548	0.0536	-0.0020	0.0137	0.0091	0.0329
4	-0.0555	0.0733	-0.0338	0.0642	0.0119	0.0652	-0.0173	0.0419	-0.0025	0.0120	-0.0099	0.0255
5	-0.0358	0.0680	-0.0677	0.0596	-0.0098	0.0613	-0.0067	0.0400	-0.0048	0.0105	0.0175	0.0266
6	-0.0411	0.0638	-0.0377	0.0582	-0.0065	0.0571	-0.0109	0.0355	0.0090	0.0122	0.0166	0.0240
7	-0.0243	0.0620	-0.0961	0.0571	0.0335	0.0551	-0.0051	0.0343	0.0044	0.0103	0.0039	0.0218
8	-0.0531	0.0622	-0.0625	0.0577	0.0669	0.0562	-0.0257	0.0344	0.0010	0.0104	0.0166	0.0224
9	-0.0951	0.0626	-0.0689	0.0582	0.0833	0.0574	-0.0094	0.0346	0.0110	0.0118	0.0128	0.0225
10	-0.1312	0.0622	-0.0847	0.0586	0.1696	0.0598	-0.0038	0.0358	0.0130	0.0132	0.0048	0.0225
11	-0.1350	0.0627	-0.0770	0.0595	0.1658	0.0610	-0.0256	0.0366	0.0222	0.0138	-0.0056	0.0220
12	-0.1509	0.0622	-0.0901	0.0587	0.2023	0.0607	-0.0489	0.0354	0.0278	0.0152	-0.0081	0.0219
Wealth	0.0077	0.0054	-0.0067	0.0053	0.0112	0.0074	-0.0068	0.0037	-0.0018	0.0023	0.0041	0.0033
Constant	0.2149	0.1013	-0.7568	0.0791	-0.2023	0.0930	0.4163	0.0769	0.0327	0.0219	-0.0295	0.0319

Notes: Estimation results are obtained from the questionnaire sample.

Table A.20: College choice preliminary estimation results

	1		2	
	Big Cities		Same city	
Variables	Coef	St Err	Coef	St Err
College Entrance Exam Score	0.003	0.000	0.001	0.000
Scaled High School GPA	0.000	0.000	0.000	0.000
Female	0.010	0.011	0.022	0.013
		Father Education Level		
Elementary School	-0.004	0.016	-0.035	0.019
Middle School	-0.009	0.022	-0.046	0.025
High School	-0.023	0.018	-0.025	0.024
Higher Education	0.102	0.044	0.043	0.066
		Income categories		
2	0.092	0.044	-0.028	0.054
3	0.030	0.035	-0.038	0.049
4	0.005	0.032	0.037	0.058
5	0.053	0.034	-0.020	0.048
6	0.021	0.030	-0.007	0.047
7	0.038	0.030	-0.008	0.046
8	0.042	0.031	-0.004	0.046
9	0.042	0.034	0.009	0.050
10	0.092	0.040	0.049	0.052
11	0.169	0.049	0.045	0.058
12	0.198	0.052	0.069	0.056
		Wealth categories		
1	0.068	0.059	0.114	0.068
2	0.036	0.054	0.107	0.042
3	0.054	0.053	0.084	0.036
4	0.053	0.053	0.065	0.036
5	0.074	0.053	0.076	0.038
6	0.078	0.054	0.068	0.039
Mother's Occupation categories	*Yes*		*Yes*	
Father's Occupation categories	*Yes*		*Yes*	
Constant	-0.720	0.091	-0.143	0.099
Obs	*1491*		*1491*	

Notes: Estimation results are obtained from the questionnaire sample and students who are away from big cities. This table presents preliminary estimation results obtained by two separate OLS regressions. Students' choice indicators are regressed on socioeconomic characteristics.

are opposite. Also there are considerable differences between same city and distance preferences between male and female students.

Table A.21: Summary statistics by gender

	Mean	SD	Min	Max	Obs
Male					
College Exam	0.77	0.09	0.62	1	106086
HS Exam	0.70	0.10	0.39	1	65535
HS GPA	70.97	23.77	0	100	30919
Medical	0.04	0.17	0	1	106086
Engineering	0.32	0.35	0	1	106086
Math	0.05	0.16	0	1	106086
Private College	0.08	0.22	0	1	104174
Same city	0.17	0.27	0	1	106086
Distance	513.55	362.85	0	2500	105961
Female					
College Exam	0.76	0.09	0.62	1	66865
HS Exam	0.70	0.10	0.40	1	46520
HS GPA	81.61	18.30	0	100	22122
Medical	0.04	0.17	0	1	66865
Engineering	0.14	0.23	0	1	66865
Math	0.05	0.16	0	1	66865
Private College	0.06	0.20	0	1	66028
Same city	0.21	0.30	0	1	66865
Distance	413.20	285.19	0	2500	66753

A.5 The effects of preference submission time on cutoffs

Lien et al. (2016), Lien et al. (2017), Wu and Zhong (2014) analyze preference submission time effect on college-student matching outcomes and they show that before exam submission with Boston mechanism produce ex-ante fair allocation compare to submission after score announcement. Inspiring from these papers, I control submission timing effect on college cutoff scores in Turkey's college-student matching.

Table A.22: Cutoff scores

Cutoff 1998			Cutoff 1999			Cutoff 2000		
Variables	Coef	Std Err	Variables	Coef	Std Err	Variables	Coef	Std Err
Cutoff 97	0.94	0.01	Cutoff 98	0.95	0.01	Cutoff 99	0.91	0.01
Quota 98	0.05	0.02	Quota 98	0.00	0.00	Quota 2000	0.00	0.00
Quota 97	-0.04	0.01	Quota 99	0.00	0.00	Quota 99	0.00	0.00
Bogazici	1.98	1.10	Bogazici	0.02	0.01	Bogazici	0.02	0.00
Itu	6.58	0.77	Itu	0.07	0.00	Itu	0.00	0.00
Odtu	4.39	0.95	Odtu	0.04	0.01	Odtu	0.01	0.00
Koc	-2.59	1.96	Koc	-0.02	0.01	Koc	0.00	0.01
Bilkent	4.42	3.10	Bilkent	-0.05	0.02	Bilkent	0.03	0.02
Ytu	8.58	0.79	Ytu	0.05	0.02	Ytu	0.00	0.00
Hacettepe	5.37	1.39	Hacettepe	-0.01	0.01	Hacettepe	-0.01	0.00
Marmara	0.78	0.69	Marmara	0.02	0.01	Marmara	0.01	0.01
Kultur			Kultur	0.05	0.02	Kultur	-0.01	0.01
Fatih	6.20	2.90	Fatih	-0.10	0.03	Fatih	0.00	0.02
Yeditepe	11.57	2.85	Yeditepe	-0.03	0.02	Yeditepe	0.01	0.01
Bilgi	20.50	4.31	Bilgi	-0.03	0.01	Bilgi	-0.02	0.01
Yuzuncuyil	-0.79	1.47	Yuzuncuyil	0.00	0.02	Yuzuncuyil	0.02	0.01
Ataturk	-1.55	1.10	Ataturk	-0.03	0.01	Ataturk	0.01	0.00
Constant	19.03	2.60	Constant	-0.05	0.01	Constant	0.05	0.00

Table A.22 shows regression results of cutoff scores of 1998, 1999, 2000 on one lagged year's cutoff scores, quotas and selected college dummies. The policy change in the preference submission time occurred in 1999, therefore 1999 cutoff scores are the result of new submission policy. It is expected to see high correlation between these cutoff scores across years but change in the college fixed effects show either change in the preference in society or change in the submission strategies. The result table shows that private colleges cutoff scores are negatively affected from submission time policies.

A.6 **Econometric framework for Chapter 2**

Following Wooldridge (2014) section 4, I construct joint QMLE setting. The setup for the fractional response variable f, endogeneity of covariate as an omitted variable problem and start by assuming p has a linear reduced form with the substantive restrictions:

$$E(f|p,z,\eta) = E(f|p,z_1,\eta) = \Phi(x_1\beta+\eta) \tag{A.9}$$

$$p = z\gamma + \varepsilon \tag{A.10}$$

where x_1 is a general function of p, z_1 and η is an omitted factor thought to be correlated with p. The first equality imposes at least one exclusion restriction, where a strict subset z_1 of z appears in $E(f|p,z,\eta)$. I can take p to be the function of endogenous explanatory variable(EEV) so that an additive, independent error ε. In the second equality which is in additive form, I assume that (η,ε) independent of z and jointly normal.

The fractional response with EEV leads the same log likelihood as the probit model specification of f, even though f is a fractional response. In order to see this, consider the average structural function under normality assumption, $\Phi(\tilde{x}_1\beta/(1+\sigma_\eta^2)^{1/2})$ where $\tilde{x}_1 = (x_1,\tilde{p})$ and Φ denotes standard normal CDF, and I hope to estimate $\beta_\eta \equiv \beta/(1+\sigma_\eta^2)^{1/2}$. Note that, I can obtain $E(f|x,\tilde{p},\eta)$ as $E(f|x,\tilde{p},\eta) = E\{1[\tilde{x}_1\beta+\eta+r \geq 0]|x,\tilde{p},\eta\}$ where distribution of r given $x,\tilde{p},\eta$, $D(r|x,\tilde{p},\eta) = Normal(0,1)$. By iterated expectations,

$$E(f|x,\tilde{p}) = E\{1[\tilde{x}_1\beta+\eta+r \geq 0]|x,\tilde{p}\} \tag{A.11}$$

$$= E\{1[\tilde{x}_1\beta_\eta + e \geq 0]|x,\tilde{p}\} \tag{A.12}$$

where $e = (\eta+r)/(1+\sigma_\eta^2)^{1/2}$ is independent of z with standard normal distribution. In order to identify this specification, I assume joint normality of (e,ε) variables which are independent of

x. Let $\rho = Corr(e, \varepsilon)$. From the equation A.12, $E(f|x,\tilde{p})$ has exactly the form of $P(w=1|x,\tilde{p})$, where $w = 1[\tilde{x}_1\beta_\eta + e \geq 0]$. In this framework, even though f is not binary, its expected value given $x, \tilde{p}$ is the same as the response probability implied by the bivariate probit model. Under the distributional assumption, I can estimate college entrance exam score formation and choice equations using joint QMLE methods with log-likelihood function

$$\max_{\theta_1,\theta_2} \sum_{i=1}^{N} [q_1(f_i, \tilde{p}_i, z_i, \theta_1, \theta_2) + q_2(\tilde{p}_i, z_i, \theta_2)]$$

where $\theta_1 = (\alpha_\eta, \beta'_{1\eta}, \beta_{2\eta}, \rho)'$ and $\theta_2 = (\delta, \gamma', \tau^2)'$ [25]

$$\begin{aligned} q_1(f,p,x,\theta_1,\theta_2) =& f\ \{log\ \Phi[\frac{\alpha_\eta + x_1\beta_{1\eta} + \tilde{p}\beta_{2\eta} + (\rho/\tau)(\tilde{p} - \delta - z\gamma)}{(1-\rho^2)^{1/2}}]\} \\ &+ (1-f)\ log\ \{1 - \Phi[\frac{\alpha_\eta + x_1\beta_{1\eta} + \tilde{p}\beta_{2\eta} + (\rho/\tau)(\tilde{p} - \delta - x\gamma)}{(1-\rho^2)^{1/2}}]\} \\ q_2(p,x,\theta_2) =& -log(\tau^2)/2 - (\tilde{p} - \delta - x\gamma)^2/(2\tau^2) \end{aligned}$$

A.6.1 Construction fractional response variables

Tables A.23 and A.24 present 3 different ROLs and their transformations according to a specific college character. For this example, I choose private college is the college type to construct college contest variable. The colleges are in the ROLs divided into private and public colleges. We use this information in ROLs transforming them binary variables such that it takes 1 if the college in ROLs is private. According to this transformations, the constructed contest variables become 0.43, 1, 0 respectively.

[25] The parameters in the fractional response variable equations are identified up to scale. $\alpha_\eta, \beta'_{1\eta}, \beta_{2\eta}$ represents scaled versions of $\alpha, \beta'_1, \beta_2$. I refer the reader to Wooldridge (2014) for detailed identification and estimation under these conditions.

Table A.23: Example ROLs

	ROL_1	ROL_2	ROL_3
1	Boun	Bilgi	Odtu
2	Boun	Bilgi	Odtu
3	Boun	Yeditepe	Gazi
4	Koc		Hacettepe
5	Sabanci		Hacettepe
6	Bilkent		Ankara
7	Odtu		Ankara
8	Odtu		Ankara
9	Sabanci		
10	Koc		
11	Itu		
12	Itu		
13	Itu		
14	itu		
15	Bilgi		
16	Bilgi		
17	Bilgi		
18	Ytu		
19	Ytu		
20	Yeditepe		
21	Yeditepe		
22	Marmara		
23	Marmara		
24	Istanbul		

Table A.24: Transformed ROLs

	ROL_1	ROL_2	ROL_3
1	0	1	0
2	0	1	0
3	0	1	0
4	1		0
5	1		0
6	1		0
7	0		0
8	0		0
9	1		
10	1		
11	0		
12	0		
13	0		
14	0		
15	1		
16	1		
17	1		
18	0		
19	0		
20	1		
21	1		
22	0		
23	0		
24	0		

A.6.2 **Additional summary and estimation results tables**

Table A.25: Summary statistics for students that are in the college admission process

Variables	Mean	SD	Obs
Panel A: Student Characteristics			
Female	0.43	0.50	1851618
Without Exam	0.10	0.30	1851618
Without Score	0.13	0.33	1851618
Panel B: Choice Variables			
Submission	0.48	0.50	1852076
# Choice	6.20	8.65	1851618
Matched	0.33	0.47	1852076
Panel C: Education tracks			
Quantitative	0.37	0.48	1844714
Verbal	0.21	0.41	1844714
Composite	0.39	0.49	1844714
Foreign Language	0.02	0.15	1844714

Table A.26: Summary statistics for college applicants from survey sample

	Mean	SD	Min	Max	Obs
Panel A: Student Characteristics					
Female	0.48	0.50	0	1	4516
Father Educ.	2.12	1.19	1	5	4442
Mother Educ.	2.80	1.28	1	5	4467
Income	7.82	2.50	1	12	4195
Wealth	4.84	1.28	0	6	4147
Panel B: College Exam Scores					
Quantitative	0.72	0.15	0	1.00	4516
Verbal	0.77	0.10	0	0.98	4516
Composite	0.77	0.10	0	0.99	4516
Panel C: High school Scores					
Quantitative	0.74	0.10	0.45	1.00	4075
Verbal	0.76	0.10	0.48	1.00	4075
Scaled GPA	0.74	0.09	0.42	1.00	4515

Table A.27: QMLE result for private college choice without endogenous score

Variables	Coef	St Err
Panel A. Covariates		
College Entrance Score	-2.33	0.06
Female	-0.21	0.01
Constant	0.65	0.15
Panel B. High school type fixed effects		
Anatolian (S)	-0.03	0.14
Anatolian Vocational (S)	-0.79	0.15
Anatolian Teacher (S)	-0.35	0.15
Regional (N)	-0.80	0.15
Science (S)	0.53	0.14
Vocational (N)	-1.05	0.15
Private Science (S)	0.68	0.14
Private (N)	0.74	0.14
Regional (N)	-0.50	0.14
Regional (S)	-0.40	0.14
City Dummies	Yes	
Obs	110,807	

Table A.28: QMLE result for private college choice with endogenous college entrance exam score

Private College Choice			College Entrance Exam Score		
Variables	Coef	St Err	Variables	Coef	St Err
Covariates			*Covariates*		
College Exam Score	1.63	0.20	High School Exam Score	0.48	0.00
Female	-0.12	0.02	Female	-0.01	0.00
Constant	-2.54	0.32	Constant	0.47	0.01
High school type fixed effects			*High school type fixed effects*		
Anatolian (S)	-0.24	0.28	Anatolian (S)	0.02	0.01
Anatolian Vocational (S)	-0.45	0.29	Anatolian Vocational (S)	-0.06	0.01
Anatolian Teacher (S)	-0.50	0.28	Anatolian Teacher (S)	0.00	0.01
Regional (N)	-0.55	0.28	Regional (N)	-0.02	0.01
Science (S)	0.00	0.28	Science (S)	0.03	0.01
Vocational (N)	-0.64	0.29	Vocational (N)	-0.04	0.01
Private Science (S)	0.67	0.28	Private Science (S)	-0.02	0.01
Private (N)	0.58	0.28	Private (N)	0.00	0.01
Regional (N)	-0.30	0.28	Regional (N)	-0.02	0.01
Regional (S)	-0.38	0.28	Regional (S)	0.01	0.01
City Dummies	Yes		City Dummies	Yes	
ρ	-0.34	0.03	Obs	110,807	
τ	0.07	0.00			

Notes: Standard errors are calculated from bootstrap simulations with 100 replications.

Table A.29: QMLE result for private college choice with endogenous high school GPA

Private College Choice			Scaled high school GPA formation		
Variables	Coef	St Err	Variables	Coef	St Err
Scaled high school GPA	1.36	0.10	High school exam score	0.70	0.01
Female	-0.29	0.02	Female	0.11	0.00
Constant	-1.29	0.20	Constant	-0.06	0.02
High school type fixed effects			*High school type fixed effects*		
Anatolian (S)	-0.56	0.19	Anatolian (S)	0.14	0.02
Anatolian Vocational (S)	-1.02	0.20	Anatolian Vocational (S)	0.22	0.02
Anatolian Teacher (S)	-0.98	0.19	Anatolian Teacher (S)	0.14	0.02
Regional (N)	-1.28	0.22	Regional (N)	0.25	0.02
Science (S)	-0.38	0.19	Science (S)	0.06	0.02
Vocational (N)	-1.24	0.21	Vocational (N)	0.25	0.03
Private Science (S)	0.30	0.19	Private Science (S)	0.13	0.02
Private (N)	0.41	0.19	Private (N)	0.02	0.02
Regional (N)	-0.63	0.19	Regional (N)	0.17	0.02
Regional (S)	-0.70	0.19	Regional (S)	0.18	0.02
ρ	-0.45	0.02	Obs	45,722	
$\exp(\tau)$	-1.59	0.00			

Notes: Standard errors are calculated from bootstrap simulations with 100 replications.

Table A.30: QMLE result for major choice without endogenous college entrance exam score

	Medical School		Engineering		Math	
Variables	Coef	St Err	Coef	St Err	Coef	St Err
College Exam Score	14.31	0.22	1.39	0.04	4.17	0.05
Female	0.23	0.01	-0.70	0.01	0.10	0.01
Constant	-14.91	0.27	-0.97	0.10	-5.26	0.11
			High school type fixed effects			
Anatolian (S)	0.64	0.18	0.36	0.10	0.02	0.11
Anatolian Vocational (S)	-0.19	0.29	-0.78	0.10	-0.99	0.16
Anatolian Teacher (S)	0.72	0.19	-0.48	0.10	-0.65	0.11
Regional (N)	1.01	0.19	0.07	0.10	0.37	0.11
Science (S)	0.75	0.19	0.17	0.10	-0.63	0.11
Vocational (N)	0.58	0.25	-1.04	0.10	-0.66	0.16
Private Science (S)	0.82	0.19	0.47	0.10	0.06	0.11
Private (N)	0.51	0.19	0.52	0.10	-0.22	0.11
Regional (N)	0.74	0.19	0.16	0.10	0.37	0.11
Regional (S)	0.61	0.19	0.29	0.10	0.34	0.11

Notes: Standard errors are calculated from bootstrap simulations with 100 replications.

Table A.31: QMLE results for major choice with endogenous college entrance exam score

	Medical School		Engineering		Math	
Variables	Coef	St Err	Coef	St Err	Coef	St Err
College Exam Score	13.21	0.24	3.26	0.09	0.09	0.14
Female	0.18	0.01	-0.65	0.01	0.02	0.01
Constant	-14.10	0.30	-2.45	0.11	-1.88	0.16
			High school type fixed effects			
Anatolian (S)	0.83	0.20	0.19	0.08	0.26	0.13
Anatolian Vocational (S)	-0.11	0.31	-0.67	0.10	-1.28	0.20
Anatolian Teacher (S)	0.89	0.20	-0.60	0.08	-0.38	0.14
Regional (N)	1.01	0.21	0.05	0.09	0.24	0.13
Science (S)	1.03	0.20	-0.14	0.09	-0.08	0.14
Vocational (N)	0.57	0.27	-0.89	0.08	-0.94	0.17
Private Science (S)	0.91	0.20	0.44	0.09	0.08	0.13
Private (N)	0.76	0.22	0.38	0.09	-0.03	0.13
Regional (N)	0.78	0.20	0.19	0.08	0.19	0.13
Regional (S)	0.76	0.20	0.23	0.08	0.34	0.13
ρ	0.11	0.01	-0.17	0.01	0.35	0.01
$\exp(\tau)$	-2.61	0.00	-2.61	0.00	-2.61	0.00

Notes: Standard errors are calculated from bootstrap simulations with 100 replications.

Table A.32: Cutoff percentiles summaries by public/private colleges and engineering majors

		Public College	
		No	Yes
Engineering	No	44.56 (33.89)	65.47 (24.98)
	Yes	54.91 (30.38)	86.41 (12.10)

Notes: Table presents mean and standard deviations (in parentheses) of percentiles according to private/public and engineering major or not partitions.

Table A.33: Variance decomposition in the ROLs

Selected Choice Variables	R^2
Fraction of private colleges	18%
Fraction of engineering major	20%
Fraction of medical schools	22%
Fraction of colleges from the same city	36%
Mean distance	27%

Notes: This table show the explained variation by admissions scores, gender, high school type, and city dummies for the selected choice variables.

Table A.34: QMLE result for same city choice with endogenous college entrance exam score

Private College Choice			College Entrance Exam Score		
Variables	Coef	St Err	Variables	Coef	St Err
College Entrance Score	1.06	0.38	High School Exam Score	0.44	0.01
Female	0.17	0.02	Female	-0.02	0.00
Constant	-1.37	0.43	Constant	0.50	0.01
High school type fixed effects			*High school type fixed effects*		
Anatolian (S)	-0.89	0.31	Anatolian (S)	0.02	0.01
Anatolian Vocational (S)	-0.64	0.32	Anatolian Vocational (S)	-0.07	0.01
Anatolian Teacher (S)	-0.96	0.32	Anatolian Teacher (S)	0.00	0.01
Regional (N)	-0.72	0.32	Regional (N)	-0.02	0.01
Science (S)	-1.09	0.32	Science (S)	0.04	0.01
Vocational (N)	-0.72	0.33	Vocational (N)	-0.05	0.01
Private Science (S)	-0.54	0.31	Private Science (S)	-0.01	0.01
Private (N)	-0.55	0.34	Private (N)	-0.01	0.01
Regional (N)	-0.50	0.31	Regional (N)	-0.01	0.01
Regional (S)	-0.62	0.31	Regional (S)	0.01	0.01
ρ	-0.15	0.03	Obs	37150	
$\exp(\tau)$	-2.62	0.00			

Notes: Standard errors are calculated from bootstrap simulations with 100 replications.

Table A.35: QMLE result for private college choice with endogenous college entrance exam score from the survey sample

Private College Choice			College Entrance Exam Score		
Variables	Coef	St Err	Variables	Coef	St Err
Constant	-4.82	0.52	Constant	0.20	0.05
College Exam Score	-0.67	0.36	High School Exam Score	0.93	0.02
Female	-0.14	0.05	Female	-0.03	0.00
Father Education			*Father Education*		
Elementary School	-0.04	0.07	Elementary School	-0.01	0.00
Middle School	-0.16	0.08	Middle School	0.00	0.01
High School	-0.02	0.09	High School	-0.01	0.01
Higher Education	0.19	0.21	Higher Education	-0.03	0.02
Mother Education			*Mother Education*		
Elementary School	0.02	0.06	Elementary School	0.00	0.00
Middle School	-0.09	0.10	Middle School	0.00	0.01
High School	-0.18	0.07	High School	0.02	0.01
Higher Education	-0.29	0.18	Higher Education	0.03	0.01
Income Categories			*Income Categories*		
2	-0.36	0.82	2	0.02	0.02
3	-0.11	0.33	3	0.01	0.02
4	-0.19	0.26	4	0.02	0.02
5	-0.24	0.26	5	0.01	0.02
6	-0.13	0.22	6	0.03	0.02
7	-0.13	0.22	7	0.03	0.02
8	-0.17	0.21	8	0.04	0.01
9	-0.04	0.21	9	0.04	0.01
10	0.26	0.23	10	0.04	0.02
11	0.40	0.23	11	0.01	0.02
12	0.66	0.21	12	0.01	0.02
Wealth Categories			*Wealth Categories*		
1	4.21	0.39	1	-0.12	0.05
2	4.05	0.32	2	-0.12	0.04
3	3.96	0.30	3	-0.14	0.04
4	3.97	0.32	4	-0.14	0.04
5	4.13	0.31	5	-0.14	0.04
6	4.40	0.32	6	-0.15	0.04
High School Dummies	Yes		High School Dummies	Yes	
ρ	-0.10	0.04	Obs	3,507	
$\exp(\tau)$	-2.27	0.01			

Notes: Standard errors are calculated from bootstrap simulations with 100 replications.

Table A.36: QMLE result for evening college choice with endogenous college entrance exam score under homogenous ability assumption

Private College Choice			College Entrance Exam Score		
Variables	Coef	St Err	Variables	Coef	St Err
College Entrance Score	-12.13	0.070	Distance Ind	-0.081	0.002
Female	-0.285	0.009	Female	-0.021	0.001
Constant	9.301	0.147	Constant	0.830	0.007
High school type fixed effects			*High school type fixed effects*		
Anatolian (S)	0.648	0.127	Anatolian (S)	0.059	0.007
Anatolian Vocational (S)	-1.174	0.134	Anatolian Vocational (S)	-0.073	0.008
Anatolian Teacher (S)	0.450	0.131	Anatolian Teacher (S)	0.054	0.007
Regional (N)	-0.906	0.133	Regional (N)	-0.047	0.007
Science (S)	1.779	0.130	Science (S)	0.140	0.008
Vocational (N)	-1.431	0.138	Vocational (N)	-0.085	0.008
Private Science (S)	0.297	0.128	Private Science (S)	0.002	0.007
Private (N)	0.742	0.130	Private (N)	0.039	0.007
Regional (N)	-0.706	0.127	Regional (N)	-0.040	0.007
Regional (S)	-0.133	0.127	Regional (S)	0.005	0.007
City Dummies	Yes		City Dummies	Yes	
ρ	0.913	0.006	Obs	110,807	
$\exp(\tau)$	-2.531	0.003			

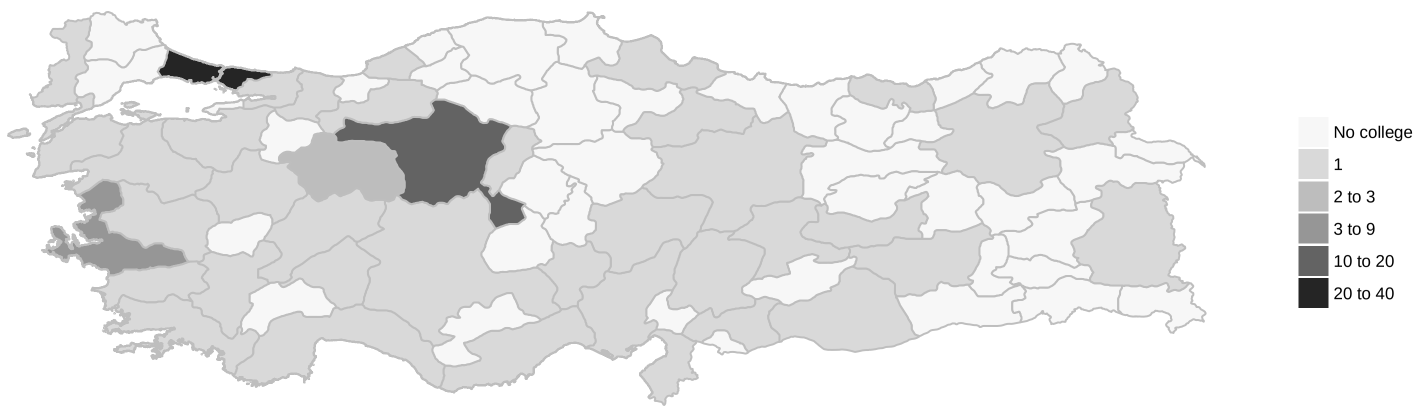
Table A.37: Linear regression of college entrance exam score

Variables	Coef.	St Err.
Panel A. Covariates		
High School Entrance Score	0.48	0.00
Female	-0.01	0.00
Constant	0.47	0.01
Panel B. High school type fixed effects		
Anatolian (S)	0.02	0.01
Anatolian Vocational (S)	-0.06	0.01
Anatolian Teacher (S)	0.00	0.01
Regional (N)	-0.02	0.01
Science (S)	0.03	0.01
Vocational (N)	-0.04	0.01
Private Science (S)	-0.02	0.01
Private (N)	0.00	0.01
Regional (N)	-0.02	0.01
Regional (S)	0.01	0.01
City Dummies	Yes	

Table A.38: QMLE result for evening college choice with endogenous college entrance exam score

Evening Education Choice			College Entrance Exam Score		
Variables	Coef	St Err	Variables	Coef	St Err
College Entrance Score	1.63	0.20	High School Exam Score	0.48	0.00
Female	-0.12	0.02	Female	-0.01	0.00
Constant	-2.54	0.32	Constant	0.47	0.01
High school type fixed effects			*High school type fixed effects*		
Anatolian (S)	-0.24	0.28	Anatolian (S)	0.02	0.01
Anatolian Vocational (S)	-0.45	0.29	Anatolian Vocational (S)	-0.06	0.01
Anatolian Teacher (S)	-0.50	0.28	Anatolian Teacher (S)	0.00	0.01
Regional (N)	-0.55	0.28	Regional (N)	-0.02	0.01
Science (S)	0.00	0.28	Science (S)	0.03	0.01
Vocational (N)	-0.64	0.29	Vocational (N)	-0.04	0.01
Private Science (S)	0.67	0.28	Private Science (S)	-0.02	0.01
Private (N)	0.58	0.28	Private (N)	0.00	0.01
Regional (N)	-0.30	0.28	Regional (N)	-0.02	0.01
Regional (S)	-0.38	0.28	Regional (S)	0.01	0.01
City Dummies	Yes		City Dummies	Yes	
ρ	-0.34	0.07	Obs	110,807	
$\exp(\tau)$	-2.62	0.00			

Figure A.7: Number of colleges by cities



A.7 Additional tables and figures for Chapter 3

Table A.39: Regression results for Earnings

Dependent	Earnings	
	Coef	Std Err
Female	-0.27	0.07
Birth Year	-0.14	0.02
Working hours	0.10	0.01
Hispanic	0.04	0.10
Black	-0.08	0.09
Two Year	-0.16	0.10
Grad School	-0.49	0.31
Rural	0.11	0.07
West	0.18	0.11
South	0.24	0.09
North East	0.42	0.10
Overall GPA	0.33	0.04
Education	-0.16	0.14
Business	0.22	0.10
Health	0.29	0.13
Computer& Eng	0.62	0.14
Other	0.04	0.09

Notes: OLS regressions result from college graduates' earnings

Table A.40: Distribution of college-majors in the NLYS97 sample

	Female		Male		Total	
Majors	Nr	(%)	Nr	(%)	Nr	(%)
Agriculture, Agriculture Operations	3	0.37	4	0.57	7	0.46
Natural Resources and Conservation	2	0.25	7	1	9	0.6
Architecture and Related Services	3	0.37	5	0.72	8	0.53
Area, Ethnic, Cultural, Gender	5	0.61	3	0.43	8	0.53
Communication, Journalism	28	3.44	23	3.3	51	3.38
Communications Technologies	1	0.12	6	0.86	7	0.46
Computer and Information Sciences	14	1.72	55	7.89	69	4.57
Personal and Culinary Services	8	0.98	6	0.86	14	0.93
Education	81	9.95	21	3.01	102	6.75
Engineering	12	1.47	42	6.03	54	3.57
Engineering Technologies	6	0.74	41	5.88	47	3.11
Foreign Languages, Literatures	12	1.47	4	0.57	16	1.06
Family and Consumer Sciences	13	1.6	.	.	13	0.86
Legal Professions and Studies	3	0.37	1	0.14	4	0.26
English Language and Literature/Letters	16	1.97	6	0.86	22	1.46
Liberal Arts and Sciences, Humanities	89	10.93	60	8.61	149	9.86
Biological and Biomedical Sciences	25	3.07	15	2.15	40	2.65
Mathematics and Statistics	3	0.37	11	1.58	14	0.93
Military Science, Leadership	1	0.12	1	0.14	2	0.13
Military Technologies and Applied Sciences	.	.	1	0.14	1	0.07
Multi/Interdisciplinary Studies	8	0.98	14	2.01	22	1.46
Parks, Recreation and Leisure Studies	9	1.11	11	1.58	20	1.32
Leisure and Recreational Activities	1	0.12	.	.	1	0.07
Philosophy and Religious Studies	2	0.25	7	1	9	0.6
Theology and Religious Vocations	5	0.61	9	1.29	14	0.93
Physical Sciences	7	0.86	7	1	14	0.93
Science and Technologies/Technicians	1	0.12	.	.	1	0.07
Psychology	56	6.88	15	2.15	71	4.7
Homeland Security, Law Enforcement	27	3.32	32	4.59	59	3.9
Public Administration and Social Service	23	2.83	4	0.57	27	1.79
Social Sciences	47	5.77	37	5.31	84	5.56
Construction Trades	.	.	7	1	7	0.46
Mechanic and Repair Technologies	.	.	16	2.3	16	1.06
Precision Production	.	.	3	0.43	3	0.2
Transportation and Materials Moving	2	0.25	6	0.86	8	0.53
Visual and Performing Arts	40	4.91	36	5.16	76	5.03
Health Professions	112	13.76	26	3.73	138	9.13
Business, Management, Marketing	133	16.34	141	20.23	274	18.13
History	13	1.6	13	1.87	26	1.72
Other	3	0.37	1	0.14	4	0.26
Total	814	100	697	100	1511	100

Table A.41: Marital status and marriage time by college-major

	Marriage Dummy			Marriage Time		
Majors	Female	Male	Total	Female	Male	Total
Agriculture, Agriculture Operations	1.00	0.75	0.86	200.33	172.50	184.43
Natural Resources and Conservation	0.00	0.29	0.22	234.00	217.14	220.89
Architecture and Related Services	0.67	0.80	0.75	177.67	190.40	185.63
Area, Ethnic, Cultural, Gender, and Group Studies	0.60	0.33	0.50	168.00	216.33	186.13
Communication, Journalism	0.57	0.52	0.55	198.46	205.39	201.59
Communications Technologies/Technicians	0.00	0.50	0.43	234.00	199.33	204.29
Computer and Information Sciences	0.43	0.47	0.46	192.00	196.02	195.20
Personal and Culinary Services	0.63	0.17	0.43	178.50	212.83	193.21
Education	0.73	0.62	0.71	173.23	191.67	177.03
Engineering	0.58	0.57	0.57	186.17	190.57	189.59
Engineering Technologies	0.50	0.56	0.55	200.33	185.56	187.45
Foreign Languages, Literatures, and Linguistics	0.67	0.75	0.69	175.50	209.00	183.88
Family and Consumer Sciences/Human Sciences	0.85	.	0.85	154.15	.	154.15
Legal Professions and Studies	1.00	0.00	0.75	162.00	234.00	180.00
English Language and Literature/Letters	0.50	0.67	0.55	204.88	188.83	200.50
Liberal Arts and Sciences, Humanities	0.69	0.63	0.66	170.87	184.72	176.44
Biological and Biomedical Sciences	0.48	0.67	0.55	196.60	195.67	196.25
Mathematics and Statistics	1.00	0.64	0.71	152.00	183.18	176.50
Military Science, Leadership and Operational Art.	1.00	1.00	1.00	98.00	153.00	125.50
Military Technologies and Applied Sciences	.	1.00	1.00	.	72.00	72.00
Multi/Interdisciplinary Studies	0.50	0.57	0.55	178.25	192.79	187.50
Parks, Recreation and Leisure Studies	0.78	0.55	0.65	165.22	192.55	180.25
Leisure and Recreational Activities	1.00	.	1.00	140.00	.	140.00
Philosophy and Religious Studies	0.00	0.57	0.44	234.00	205.14	211.56
Theology and Religious Vocations	0.60	0.56	0.57	184.80	182.56	183.36
Physical Sciences	0.86	1.00	0.93	200.71	138.00	169.36
Science and Technologies/Technicians	0.00	.	0.00	234.00	.	234.00
Psychology	0.55	0.53	0.55	190.36	202.07	192.83
Homeland Security, Law Enforcement, Firefighting	0.63	0.50	0.56	172.52	188.00	180.92
Public Administration and Social Service Professions	0.39	0.75	0.44	193.04	205.75	194.93
Social Sciences	0.49	0.49	0.49	198.23	206.11	201.70
Construction Trades	.	0.57	0.57	.	193.57	193.57
Mechanic and Repair Technologies/Technicians	.	0.56	0.56	.	188.19	188.19
Precision Production	.	0.33	0.33	.	206.00	206.00
Transportation and Materials Moving	1.00	0.67	0.75	213.50	177.83	186.75
Visual and Performing Arts	0.50	0.56	0.53	185.30	194.42	189.62
Health Professions and Related Programs	0.56	0.58	0.57	187.02	195.88	188.69
Business, Management, Marketing	0.63	0.50	0.56	176.80	195.79	186.57
History	0.31	0.38	0.35	213.85	205.38	209.62
Other	0.00	1.00	0.25	234.00	118.00	205.00
Total	0.60	0.55	0.57	183.42	193.38	188.01

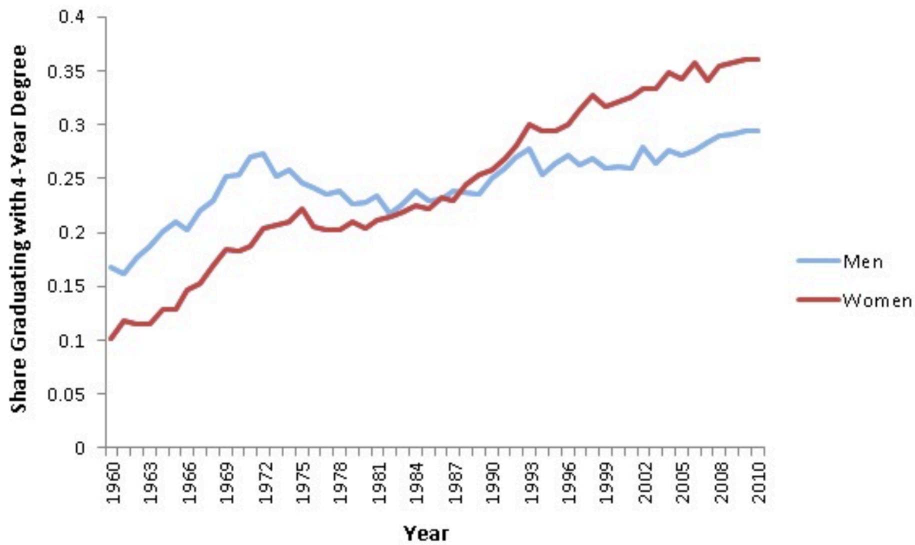
Table A.42: College-major graduates numbers and fractions in the population by the years

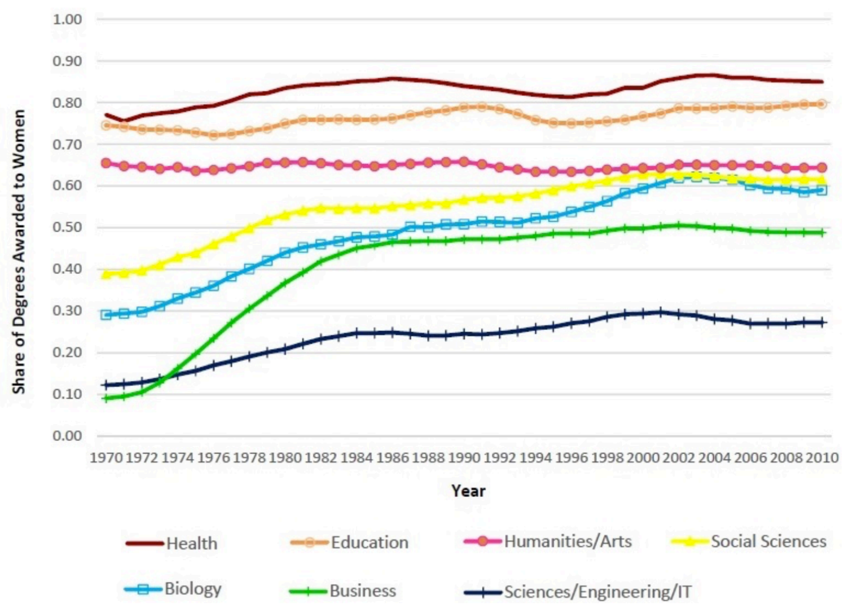
Field of study	2003-04		2004-05		2005-06		2006-07		2007-08		2008-09		2009-10		2010-11		2011-12		2012-13		2013-14	
Agriculture	22,835	1.6%	23,002	1.6%	23,053	1.6%	23,133	1.5%	24,113	1.5%	24,982	1.6%	26,343	1.6%	28,630	1.7%	30,972	1.7%	33,592	1.8%	35,116	1.9%
Architecture	8,838	0.6%	9,237	0.6%	9,515	0.6%	9,717	0.6%	9,805	0.6%	10,119	0.6%	10,051	0.6%	9,831	0.6%	9,727	0.5%	9,757	0.5%	9,144	0.5%
Area. ethnic. cultural.	7,181	0.5%	7,569	0.5%	7,879	0.5%	8,194	0.5%	8,454	0.5%	8,772	0.5%	8,620	0.5%	8,955	0.5%	9,228	0.5%	8,850	0.5%	8,275	0.4%
Biological and biomedical	62,624	4.5%	65,915	4.6%	70,607	4.8%	76,832	5.0%	79,829	5.1%	82,828	5.2%	86,391	5.2%	89,984	5.2%	95,850	5.3%	100,397	5.5%	104,633	5.6%
Business	307,149	21.9%	311,574	21.6%	318,042	21.4%	327,531	21.5%	335,254	21.4%	348,056	21.7%	358,119	21.7%	365,133	21.3%	367,235	20.5%	360,887	19.6%	358,079	19.2%
Communication. Journalism	70,968	5.1%	72,715	5.1%	73,955	5.0%	74,783	4.9%	76,382	4.9%	77,984	4.9%	81,280	4.9%	83,231	4.9%	83,771	4.7%	84,818	4.6%	87,604	4.7%
Communications technologies	2,034	0.1%	2,523	0.2%	2,981	0.2%	3,637	0.2%	4,666	0.3%	37,992	2.4%	39,593	2.4%	43,066	2.5%	47,406	2.6%	50,961	2.8%	55,367	3.0%
Computer and IS	59,488	4.3%	54,111	3.8%	47,480	3.2%	42,170	2.8%	38,476	2.5%	37,992	2.4%	39,593	2.4%	43,066	2.5%	47,406	2.6%	50,961	2.8%	55,367	3.0%
Education	106,278	7.6%	105,451	7.3%	107,238	7.2%	105,641	6.9%	102,582	6.6%	101,716	6.4%	101,287	6.1%	104,008	6.1%	105,656	5.9%	104,698	5.7%	98,854	5.3%
Engineering	63,410	4.5%	64,707	4.5%	66,841	4.5%	66,874	4.4%	68,431	4.4%	68,911	4.3%	72,657	4.4%	76,356	4.4%	81,371	4.5%	85,987	4.7%	92,162	4.9%
Engineering technologies	14,669	1.0%	14,837	1.0%	14,565	1.0%	14,980	1.0%	15,177	1.0%	15,493	1.0%	16,078	1.0%	16,741	1.0%	17,283	1.0%	17,010	0.9%	16,807	0.9%
English language	53,984	3.9%	54,379	3.8%	55,096	3.7%	55,122	3.6%	55,038	3.5%	55,465	3.5%	53,229	3.2%	52,754	3.1%	53,765	3.0%	52,401	2.8%	50,404	2.7%
Family and CS	19,172	1.4%	20,074	1.4%	20,775	1.4%	21,400	1.4%	21,870	1.4%	21,906	1.4%	21,832	1.3%	22,438	1.3%	23,441	1.3%	23,930	1.3%	24,722	1.3%
Foreign languages.	17,754	1.3%	18,386	1.3%	19,410	1.3%	20,275	1.3%	20,977	1.3%	21,169	1.3%	21,507	1.3%	21,705	1.3%	21,756	1.2%	21,647	1.2%	20,335	1.1%
Health professions	80,685	5.6%	91,973	6.2%	101,810	6.7%	111,478	7.1%	120,420	7.5%	129,623	7.9%	143,463	8.4%	163,675	9.1%	181,149	9.8%	198,770	10.6%		
Homeland security.	28,175	2.0%	30,723	2.1%	35,319	2.4%	39,206	2.6%	40,235	2.6%	41,788	2.6%	43,613	2.6%	47,600	2.8%	54,091	3.0%	60,264	3.3%	62,409	3.3%
Legal professions	2,841	0.2%	3,161	0.2%	3,302	0.2%	3,596	0.2%	3,771	0.2%	3,822	0.2%	3,886	0.2%	4,429	0.3%	4,595	0.3%	4,425	0.2%	4,513	0.2%
Liberal arts and sciences.	42,106	3.0%	43,751	3.0%	44,898	3.0%	44,255	2.9%	46,940	3.0%	47,095	2.9%	46,963	2.8%	46,717	2.7%	46,961	2.6%	46,790	2.5%	45,260	2.4%
Library science	72	0.0%	76	0.0%	76	0.0%	82	0.0%	68	0.0%	78	0.0%	85	0.0%	96	0.0%	95	0.0%	102	0.0%	127	0.0%
Mathematics and statistics	13,327	1.0%	14,351	1.0%	14,770	1.0%	14,954	1.0%	15,192	1.0%	15,507	1.0%	16,029	1.0%	17,182	1.0%	18,841	1.1%	20,449	1.1%	20,980	1.1%
Military technologies	10	0.0%	40	0.0%	33	0.0%	168	0.0%	39	0.0%	55	0.0%	56	0.0%	64	0.0%	86	0.0%	105	0.0%	185	0.0%
Multi/inter studies	28,047	2.0%	28,939	2.0%	30,583	2.1%	32,111	2.1%	34,174	2.2%	35,376	2.2%	37,717	2.3%	42,473	2.5%	45,717	2.6%	47,658	2.6%	48,348	2.6%
Parks. recreation.	22,164	1.6%	22,888	1.6%	25,490	1.7%	27,430	1.8%	29,931	1.9%	31,683	2.0%	33,332	2.0%	35,934	2.1%	38,998	2.2%	42,628	2.3%	46,042	2.5%
Philosophy and religious	11,152	0.8%	11,584	0.8%	11,985	0.8%	11,969	0.8%	12,257	0.8%	12,448	0.8%	12,503	0.8%	12,830	0.7%	12,645	0.7%	12,792	0.7%	11,997	0.6%
Physical sciences and	18,131	1.3%	19,104	1.3%	20,522	1.4%	21,291	1.4%	22,179	1.4%	22,691	1.4%	23,381	1.4%	24,705	1.4%	26,664	1.5%	28,053	1.5%	29,304	1.6%
Precision production	61	0.0%	64	0.0%	55	0.0%	23	0.0%	33	0.0%	29	0.0%	29	0.0%	43	0.0%	37	0.0%	36	0.0%	37	0.0%
Psychology	82,098	5.9%	85,614	5.9%	88,134	5.9%	90,039	5.9%	92,587	5.9%	94,273	5.9%	97,215	5.9%	100,906	5.9%	109,099	6.1%	114,446	6.2%	117,298	6.3%
Public administration	20,552	1.5%	21,769	1.5%	21,986	1.5%	23,147	1.5%	23,493	1.5%	23,852	1.5%	25,421	1.5%	26,799	1.6%	29,695	1.7%	31,950	1.7%	33,483	1.8%
Social sciences and history	150,357	10.7%	156,892	10.9%	161,485	10.9%	164,183	10.8%	167,363	10.7%	168,517	10.5%	172,782	10.5%	177,169	10.3%	178,534	10.0%	177,767	9.7%	173,096	9.3%
Theology and religious	8,126	0.6%	9,284	0.6%	8,548	0.6%	8,696	0.6%	8,992	0.6%	8,940	0.6%	8,719	0.5%	9,073	0.5%	9,304	0.5%	9,385	0.5%	9,642	0.5%
Transportation	4,824	0.3%	4,904	0.3%	5,349	0.4%	5,657	0.4%	5,203	0.3%	5,189	0.3%	4,998	0.3%	4,941	0.3%	4,876	0.3%	4,661	0.3%	4,588	0.2%
Visual and performing arts	77,181	5.5%	80,955	5.6%	83,297	5.6%	85,186	5.6%	87,703	5.6%	89,143	5.6%	91,798	5.6%	93,939	5.5%	95,806	5.3%	97,799	5.3%	97,246	5.2%
Not classified by field	0	0.0%	0	0.0%	0	0.0%	0	0.0%	377	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Total	1,399,542		1,439,264		1,485,242		1,524,092		1,563,069		1,601,399		1,649,919		1,716,053		1,792,163		1,840,381		1,869,814	

Table A.43: Summary statistics for entire sample

	Female(49%)			Male(51%)			Total				
	Mean	Sd	Obs	Mean	Sd	Obs	Mean	Sd	Obs	Min	Max
Birth month	30.55	17.18	4385	30.80	17.01	4599	30.68	17.09	8984	1	60
Marital Status	0.49	0.50	4385	0.40	0.49	4599	0.45	0.50	8984	0	1
First Marriage Month	185.98	59.21	4385	199.42	50.24	4599	192.86	55.21	8984	7	234
Two Year Ins	0.07	0.26	4385	0.05	0.21	4599	0.06	0.23	8984	0	1
Four Year Ins	0.16	0.37	4385	0.12	0.32	4599	0.14	0.35	8984	0	1
Graduate School	0.01	0.07	4385	0.00	0.07	4599	0.00	0.07	8984	0	1
Sec. Degree	0.02	0.14	4385	0.02	0.12	4599	0.02	0.13	8984	0	1
Sec. College	0.08	0.27	4385	0.05	0.21	4599	0.06	0.24	8984	0	1
North East	0.17	0.38	4385	0.18	0.39	4599	0.18	0.38	8984	0	1
North Central	0.23	0.42	4385	0.23	0.42	4599	0.23	0.42	8984	0	1
South	0.38	0.49	4385	0.37	0.48	4599	0.37	0.48	8984	0	1
West	0.22	0.42	4385	0.22	0.41	4599	0.22	0.42	8984	0	1
Black	0.27	0.44	4385	0.25	0.44	4599	0.26	0.44	8984	0	1
Hispanic	0.21	0.41	4385	0.21	0.41	4599	0.21	0.41	8984	0	1
White	0.51	0.50	4385	0.52	0.50	4599	0.52	0.50	8984	0	1
Average Earning	1.69	1.13	4111	2.21	1.53	4341	1.96	1.38	8452	0	18.03
Average working hours	12.24	5.93	4305	14.16	7.08	4508	13.22	6.61	8813	0	45.41
Gpa term1	2.47	1.31	1885	2.37	1.28	1593	2.42	1.29	3478	0	4
Gpa term2	2.44	1.29	1762	2.35	1.24	1460	2.40	1.27	3222	0	4
Gpa term3	2.53	1.25	1636	2.45	1.20	1309	2.50	1.23	2945	0	4

Figure A.8: College graduates in US across gender and majors





Source: National College Education Survey NCES (2012).The left graph shows the share graduating with a four year degree for both males and females. The right graph presents the proportion of degrees awarded to female students for the selected majors from 1970 to 2010.