Title
Essays on Econometrics and its Application to Education

Author
Sun, Xiaoting

Publication Date
2019

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Essays on Econometrics

and its Application to Education

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Economics

by

Xiaoting Sun

2019

ABSTRACT OF THE DISSERTATION

Essays on Econometrics

and its Application to Education

by

Xiaoting Sun

Doctor of Philosophy in Economics

University of California, Los Angeles, 2019

Professor Rosa Liliana Matzkin, Co-Chair

Professor Kathleen M McGarry, Co-Chair

This dissertation consists of three chapters that study econometrics questions and their applications to education. Chapter 1 studies a nonparametric two-sided many-to-one matching model, where many agents on one side match one institution on the other side. Classical examples include student-college matching and firm-worker matching. In this paper, I study nonparametric identification and estimation of many-to-one matching with non-transferable utility. The existing literature either assumes that the matching algorithm and reported preferences are observed or that preferences are homogeneous. This paper assumes heterogeneous preferences on the two sides and only requires data on who matches with whom in a single large market. Under mild restrictions, I prove that both the utility functions of the students and colleges and the joint distribution of unobserved heterogeneity from the two sides are nonparametrically identified. Based on my constructive identification results, I propose nonparametric and semiparametric estimators of the model and establish their consistency and asymptotic normality. The semiparametric estimator converges at a root-n rate.

Chapter 2 analyzes the U.S. college admissions under a many-to-one matching framework. In recovering the parameters of the utility functions, I am able to demonstrate substantial welfare consequences for different groups of students, relative to a centralized matching mechanism. In estimating the model using data from High School Longitudinal Study of 2009 (HSLS:09) and the Integrated Postsecondary Education Data System (IPEDS), I show that the students who experience the largest losses are first-generation college students and low-ability students. This potential loss among these groups provides an opportunity for policy interventions to lead to substantial gains in welfare.

Chapter 3 (joint with Kathleen McGarry) studies the three generations of changing gender patterns of schooling in China. The phenomenon of son preference in China and throughout much of Asia has been well documented. However, changing economic conditions, such as increases in educational attainment and employment opportunities for women and the rise in the prevalence of one child families, have likely changed the incentives for parents to invest in daughters. In this paper, we take advantage of data spanning three generations of Chinese families to examine the evolution of educational attainment for boys and girls and importantly the relative levels of schooling of each gender. We also use variation in the timing of compulsory schooling laws and the implementation of the one child policy to assess the effect of these policy measures on the relative educational levels. We find a substantial narrowing of the gap between the schooling of boys and girls, so much so that girls now have more schooling on average than boys. In addition, public policy initiatives had a larger effect in rural than urban areas.

The dissertation of Xiaoting Sun is approved.

Jennie Elizabeth Brand

Moshe Buchinsky

Zhipeng Liao

Shuyang Sheng

Rosa Liliana Matzkin, Committee Co-Chair

Kathleen M McGarry, Committee Co-Chair

University of California, Los Angeles

2019

DEDICATIONS

To my husband and parents

Contents

List of Figures

List of Tables

ACKNOWLEDGMENT

I am grateful to Rosa Matzkin, Kathleen McGarry, and Shuyang Sheng for invaluable guidance and continuous support. I thank Denis Chetverikov, Jinyong Hahn, Zhipeng Liao, Andres Santos, and the seminar participants of 2018 California Econometrics Conference, 2018 Midwest Econometrics Group Conference, UCLA econometrics proseminars, Auburn University, University of Florida, Rice University, Johns Hopkins University, University of Western Ontario, Simon Fraser University, National University of Singapore, and University of New South Wales for helpful comments and discussions. I thank Moshe Buchinsky and Jonathan Gu for help in obtaining the data. All errors are mine.

Chapter 3 is an original manuscript of McGarry and Sun (2018) published by Taylor & Francis in Journal of the Asia Pacific Economy, available online: `http://www.tandfonline.com/10.1080/13547860.2018.1515003`.

VITA

Education

University of California, Los Angeles	Los Angeles, USA
C.Phil., Economics, Department of Economics	2015
M.A., Economics, Department of Economics	2015
Peking University	Beijing, China
M.A., Economics, National School of Development	2013
B.A., Economics, National School of Development	2010
B.S., Management, Department of Information Management	2010

Fellowships and Awards

UCLA Dissertation Year Fellowship, 2018

UCLA Travel Grant, 2018

Marcia and Herbert Howard Graduate Fellowship (Best Paper in Econometrics), 2017

Pass with honors in Econometrics Field, 2015

UCLA Graduate University Fellowship, 2013-2017

Peking University Graduate Fellowship, 2010-2013

Best Teaching Assistant Award, 2012

Outstanding Graduate of Beijing, 2010

Undergraduate Research Fund, 2010

Leo KoGuan Scholarship, 2009

Peking University Challenge Cup Research Competition Second Prize, 2008, 2009

Tencent Scholarship, 2008

Yang Fuqing & Wang Yangyuan Scholarship, 2007

Conferences and Invited Seminars

2019 Auburn University, University of Florida, Rice University,
Johns Hopkins University, University of Western Ontario,
Simon Fraser University, National University of Singapore,
University of New South Wales

2018 Annual Meeting of Midwest Econometrics Group,
California Econometrics Conference

Teaching Experience

Instructor

Statistics for Economists, UCLA, Summer 2016

Teaching Assistant

Time Series Econometrics (Graduate Course), UCLA, Spring 2016

Statistics for Economists, UCLA, 2015 Winter, 2015 Spring, 2018 Winter, 2018 Spring

Microeconomic Theory I, UCLA, 2016 Winter

Microeconomic Theory II, UCLA, 2014 Fall, 2015 Fall

Microeconomic Theory, Peking University, 2011 Fall, 2012 Fall

Chapter 1

Identification and Estimation of Many-to-One Matching

1.1 Introduction

Two-sided many-to-one matching models study the assignment of many agents to one institution. They have broad applications in the real world — for example, in the college admissions, each college matches many students, in the labor market, each firm matches many workers. While the theoretical properties of matching models have been extensively discussed and well established, the econometric framework for recovering preferences in these markets is less well explored. The primary goal of this paper is to identify and estimate preferences in a many-to-one matching market. This includes recovering how students' and colleges' characteristics determine their preferences and how the unobserved heterogeneity is distributed. The model only requires data on matching outcomes (i.e., who matches with whom) in a single large market. Such a model is useful when the mechanism that assigns agents to institutions is not observed. I use the college admissions problem to illustrate my methodology and to demonstrate the importance of estimating preferences in a real world setting. [1]

Recovering preferences in a many-to-one matching model allows one to analyze the equilibrium outcome, evaluate the impact of alternative public policies, and quantify welfare implications of different allocation mechanisms. Specifically, this paper provides insights into the following aspects. First, the equilibrium outcome depends on the preferences of the two sides (i.e., the students and the colleges). It is thus necessary and difficult to identify the underlying preferences. Second, public policies that are designed for a particular subset of students or colleges may have equilibrium effects on all players. Third, the outcome of the matching mechanisms per se has welfare and inequality implications, many of which, including the welfare consequences of decentralized markets, are still not well understood.[2]

[1] The college admissions problem was introduced by Gale and Shapley (1962). Many matching problems fit in this model where the participants on one side of the market are individuals (e.g., students, workers, and interns) and the participants on the other side are institutions (e.g., colleges, firms, and hospitals).

[2] Examples of decentralized matching markets include college admissions in countries such as the U.S., UK, Japan and Korea, taxi market, some annual entry-level professional labor markets such as the market for new PhD economists, and so on. These markets are in contrast to centralized markets in which participants of each side submit preference rank order lists or priority rules and receive allocations from a centralized

This paper contributes to the emerging literature on the identification and estimation of matching models. Without parametrically specifying either the utility functions of students or colleges, or the joint distribution of unobserved heterogeneity of the two sides, I prove that the utility functions and the joint distribution of the unobserved terms in a many-to-one matching model are identified under mild restrictions. Following the constructive identification results, I propose estimation methods for both the nonparametric and semiparametric models. I establish the consistency and asymptotic normality of the estimator in each case. Especially, the semiparametric estimator is root-n consistent, asymptotically normal, and easy to apply. I then exploit the proposed estimator to analyze the U.S. college admissions, quantify the welfare consequences of the decentralized market relative to a centralized benchmark, and investigate channels through which the welfare can be improved.

I focus on matchings with non-transferable utility.[3] In this model, each student faces a multinomial choice set that is unobserved and determined in equilibrium. The choice sets in equilibrium essentially depend on the preferences of all colleges and students in the market; in particular, unobserved taste shocks of each student will influence her own choice set through the equilibrium. To facilitate a tractable structure, I build on the theoretical model of Azevedo and Leshno (2016), who establish a cutoff-theoretic framework for matching markets. The cutoff for any particular is defined as the minimal utility obtained from the students with whom the college is matched. Azevedo and Leshno (2016) prove that there is a one-to-one correspondence between stable matchings and market clearing cutoffs. The cutoffs, while still unobserved and endogenous in my model, serve as sufficient statistics for the characterization of the choice sets. I then use equilibrium cutoffs in the limit, where the number of students goes to infinity while the number of colleges remains fixed, as an approximation for the equilibrium cutoffs in the finite models.

clearing house. National Resident Matching Programs (NRMP) and school assignment system in New York and Boston adopt a centralized mechanism.

[3]In non-transferable utility model, there is no transfers that are negotiated by each student-college pair and endogenously determined to clear the market. For example, in school choice problem, although students pay tuition fees to schools or colleges, these monetary transfers are not determined by each matched pairs.

To identify the model, I use a set of exclusive regressors, each of which affects the preference of only one college without shifting the preference of students, to trace out the variation in the choice sets. Furthermore, because the joint distribution of students' and colleges' unobserved heterogeneity is not parametrically specified, I separate the underlying utilities from the two sides by introducing an additional set of exclusive regressors that alter only the student's preferences without changing her choice set.

The identification results consist of two parts. First, I show the identification of the derivatives of the utility functions of students and colleges. The formal proof builds on Matzkin (2018), who establishes pointwise constructive identification in a fully nonparametric discrete choice model with nonseparable errors. In Matzkin (2018), by taking derivatives of the probability of choosing an alternative outside the set, the nonparametric problem is transformed into a system of linear equations. The solution to the system of equations recovers the derivatives of the utility functions. This paper follows similar strategies on a two-sided matching model and uses the derivatives from conditional probabilities of being matched with each college. It is an extension of Matzkin (2018) by considering a discrete choice model with unobserved choice sets. Moreover, although the probability of being unmatched is sufficient for the identification, I consider the probability of being matched to each college. This setting of relatively richer information enables me to propose a root-n consistent semiparametric estimator based on the constructive identification results.

The second part of the identification results involves the joint distribution of the unobserved random terms. In classic discrete choice models, once the utility functions are identified, the distribution of the unobserved random term can be directly identified from the conditional probability of choosing the outside option. However, in a two-sided matching model, disentangling the effects of taste shocks from the two sides is challenging, especially given that the random terms from the two sides and across colleges can be correlated. The key insight here is that the outcome where a student is not matched with any college can arise for several reasons based on differences in abilities and preferences. For example, the student does not want to attend any college *and* is not qualified for any college. Or, the student is qualified for but unwilling to go to any college (possibly due to liquidity constraints). The probability of the latter scenario can be identified by "shutting down" the effects of students' qualification. The probability of the former scenario corresponds to the joint distribution of the random terms. It is identified once the probabilities of all the other scenarios such as the latter scenario are identified.

Directly following the identification results, I propose a nonparametric estimator for my model based on Kernel methods and establish its consistency and asymptotic normality. In a semiparametric model, based on my nonparametric identification results, I construct a set of moment conditions, where the elements other than the parameters of interest can be estimated using average derivative estimators. I thus propose a two-step estimator where the first step implements the average derivative estimation and the second step implements GMM estimation. I then show the root-n consistency and asymptotic normality of the semiparametric estimator.

Contributions to the Literature

This paper contributes to various strands of literature. First, this paper contributes to the literature on the identification of matching models by allowing richer heterogeneity and by separately and nonparametrically identifying utility functions from the two sides. There is an extensive literature on matching models with transferable utility (TU) (Choo and Siow, 2006; Graham, 2013; Sinha, 2015; Chiappori and Salanié, 2016; Mindruta, Moeen, and Agarwal, 2016; Fox, 2010, 2018; Fox, Yang, and Hsu, 2018, among others). Instead, this paper focuses on the small literature on matching models with nontransferable utility (NTU). While TU model is more straightforward to characterize due to the equivalence between stability and surplus maximization and the existence of unique equilibrium, it cannot separately identify utility functions from the two sides when the transfers are not observed.[4] Menzel (2015)

[4]The reason is that in a TU model, only quasi-surplus (i.e., the sum of the utilities of the matched pair) matters for the equilibrium.

establishes a simple asymptotic formula of one-to-one matching with NTU, but the model also cannot separately identify the utility functions. In contrast, in my paper, the utility functions from the two sides are separately identified, which is crucial when, for example, one is interested in the welfare of students and colleges separately. The other two papers on NTU that I am aware of allowing separate identification of the utility functions are Diamond and Agarwal (2017) and Galichon and Hsieh (2018). Diamond and Agarwal (2017) study a positive assortative matching model incorporating both TU and NTU in a nonparametric framework. The asymptotics they consider are that the number of agents on both sides goes to large. They assume that preferences are homogeneous and the attractiveness can be summarized using a single index. In comparison, my paper assumes that only the number of students goes to large and allows for fully heterogeneous preferences on the two sides. Galichon and Hsieh (2018) derive a Leontief matching function, which may be identified using multi-market data. By contrast, my paper considers one large market.

This paper is, to the best of my knowledge, the first to prove nonparametric identification of the joint distribution of unobserved heterogeneity in an NTU matching model.[5] Moreover, it allows correlation in preferences for different alternatives. Such flexibility is important because, in practice, there may be unobserved characteristics of the student that make the colleges (or the student) to exhibit correlated preferences over the student (the colleges).

By developing a framework that does not require information on matching mechanism or preference rank order lists of either students or colleges, I extend the literature on school choice using the matching framework. A few recently emerging literature on school choice (Fack, Grenet, and He, 2019; Agarwal and Somaini, 2018; Bucarey, 2018) follows and extends the cutoff-framework, assuming that the cutoffs are observed. This paper departs from the literature by assuming that the cutoffs are unobserved, which is the case in the U.S. college admissions and many labor markets.

[5] Regarding the TU model, Fox, Yang, and Hsu (2018) nonparametrically identify the distribution of unobserved Heterogeneity. They consider many-to-many matching and matching with trades using data from many markets.

My paper also relates to the growing literature on consideration set (Eliaz and Spiegler, 2011a,b; Masatlioglu, Nakajima, and Ozbay, 2012; Manzini and Mariotti, 2014, among others). In such models, agents choose the utility-maximizing alternative from a consideration set which is limited to the alternatives to which they pay attention. My model shares similar structure with these studies in the sense that agents face a choice set with some alternatives eliminated. The difference is that the limited choice set is due to the bilateral feature of a matching model (i.e., students can only choose from the colleges that admit them) rather than to limited attention. Moreover, the limited choice set in my model is determined in the equilibrium. Also, similar to these studies, my model goes beyond the Independence of irrelevant alternatives (IIA) assumption.

1.2 The Model

In this section, I study a two-sided many-to-one matching model with complete information. The students are indexed by $i \in \{1, \dots, n\}$ and colleges indexed by $j \in \{1, \dots, J\}$. Each college j's capacity relative to n is denoted by $s_j \in \mathcal{S} = (0, 1)$, which is exogenously given and observed. Assume that the total number of available seats does not exceed the total number of students in the market, i.e., $\sum_j s_j < 1$.

1.2.1 Utility

I consider a matching model with non-transferable utilities (NTU) in a single market. The preferences are represented by random utility functions of the form,

$$u_{ij} = u^j(x_i, z_{ij}, \epsilon_{ij}) \tag{1.1}$$

$$v_{ji} = v^j(z_{ij}, x_i, \eta_{ji}), \tag{1.2}$$

where u_{ij} denotes student i's preference for college j and v_{ji} denotes college j's preference for student i. Both u_{ij} and v_{ji} depend on student i's observed attribute $x_i \in \mathbb{R}^{k_x}$ and college j's observed attributes $z_{ij} \in \mathbb{R}^{k_z}$. x_i can include, for example, family background and SAT score. z_{ij} can refer to net price of attending college, which is defined as tuition minus student's scholarships or grant. The scalars ϵ_{ij} and η_{ji} are unobserved match-specific taste shocks that enter u_{ij} and v_{ji}, respectively.[6] The random terms therefore allow for heterogeneity in taste shocks over an infinite number of observed types. I emphasize that I do not impose any parametric assumptions on the joint distribution of $(\epsilon_{i1}, \ldots, \epsilon_{iJ}, \eta_{1i}, \ldots, \eta_{Ji})$ for the identification results. The taste shocks are i.i.d. across i, and for each student i, I allow the unobserved random terms to be correlated among colleges. This implies that the unobservables can nest individual-specific taste shocks.

Furthermore, assume all students are acceptable, i.e., colleges always prefer being matched to some student over keeping seats unfilled. However, students may prefer choosing some outside options over going to colleges.

1.2.2 Equilibrium

Let m_{ij} denote whether student i is matched with college j. That is, $m_{ij} = 1$ if student i is matched with college j and $m_{ij} = 0$ otherwise. A matching can be represented by $m = \{m_{ij}\}_{i=1,\ldots,n,\, j=1,\ldots,J}$. The solution concept is pairwise stability.[7]

Definition 1. A student-college **pair** (i, j) **block**s a matching m if student i prefers college j to his current match and either (i) college j still has vacant seats or (ii) there exists another student k matched with college j whose is ranked by college j below student i.

Definition 2. A matching m is **pairwise stable** if it is not blocked by any student-college

[6] The assumption of match-specific unobservables is in line with Menzel (2015) and Fox, Yang, and Hsu (2018). In contrast, Choo and Siow (2006), Fox (2010, 2018), Galichon and Salanie (2015), Sinha (2015), Galichon and Salanie (2015), and Galichon and Hsieh (2018) assume type-specific unobservables in the sense that men have identical preference for women in the same observed type.

[7] By definition, pairwise stable matchings are robust to one-link deviations. It is a weak equilibrium notation that is very commonly used in a matching context.

pair.

Given a stable matching m, define the associated cutoff of college j by the lowest utility it obtains from the students matched with it:

$$p_j = \inf_{\{i:\, m_{ij}=1\}} v_{ji}.$$

Let $p = (p_1, \ldots, p_J)^{'}$. There is a one-to-one correspondence between stable matchings and market clearing cutoffs (Azevedo and Leshno, 2016). I thus can summarize a stable matching using a $J \times 1$ vector p. Throughout the paper, I assume that p is unobserved.

We observe matching outcome m, college's capacities $(s_1, \ldots, s_J)$, and i.i.d. data on characteristics of students and colleges $\{x_i, z_{ij}\}_{i=1}^{n}$. Let ϵ_i and η_i denote the vectors of unobserved random shocks of student i's preferences over colleges and colleges' preferences over student i, respectively, i.e.,$\epsilon_i = (\epsilon_{i1,} \ldots, \epsilon_{iJ})^{'}, \eta_i = (\eta_{1i,} \ldots, \eta_{Ji})^{'}$. Let $z_i = (z_{i1}^{'}, \ldots, z_{iJ}^{'})^{'}$and $z_{i,-j} = (z_{i1}^{'}, \ldots, z_{ij-1}, z_{ij+1}, \ldots, z_{iJ}^{'})^{'}$. For a given vector p, the choice set of student i can be written as

$$C(x_i, z_i, \eta_i; p) = \{j:\, p_j \leq v^j(z_{ij}, x_i, \eta_{ji})\},$$

Given the choice set, student i make the decision of whether to be matched with college j as a "price-taker"

$$\begin{aligned}
& m_{ij}(x_i, \epsilon_i, \eta_i, z_{ij}; z_{i,-j}, p) \\
=& 1\Big(j = \underset{\{j^{'} \in Ch(x_i, z_i, \eta_i; p)\}}{\arg\max} u^{j^{'}}(x_i, z_{j^{'}}, \epsilon_{ij^{'}})\Big) \\
=& \underbrace{1(j \in C(x_i, z_i, \eta_i; p)}_{\text{Admitted by college } j} \\
& \times \prod_{j^{'} \neq j} \Big[1 - \underbrace{1\Big(u^j(x_i, z_{ij}, \epsilon_{ij}) < u^{j^{'}}(x_i, z_{ij^{'}}, \epsilon_{ij^{'}})\Big)}_{\text{college } j^{'} \succ \text{college } j} \underbrace{1(j^{'} \in C(x_i, z_i, \eta_i; p)}_{\text{Admitted by college } j^{'}}\Big]
\end{aligned}$$

The above expression implies that student i is matched with college j if and only if (i) college j is available to student i, and (ii) there does not exist any other college $j^{'}$ who forms a blocking pair with student i (i.e., there does not exist any other college $j^{'}$ that is feasible to student i and will give student i higher utility than college j). Given the cutoff, the "decision" of being matched with college j depends on all the colleges' characteristics $(z_{i1}, \ldots, z_{iJ})$ in addition to student i's own characteristics $(x_i, \eta_i, \epsilon_i)$.

Let $x^n = (x_1^{'}, \ldots, x_n^{'})^{'}$, $z_j^n = (z_{1j}^{'}, \ldots, z_{nj}^{'})^{'}$, $z_{-j}^n = (z_{1,-j}^{'}, \ldots, z_{n,-j}^{'})^{'}$, $z^n = (z_j^n, z_{-j}^n)$, $\epsilon^n = (\epsilon_1^{'}, \ldots, \epsilon_n^{'})^{'}$, and $\eta^n = (\eta_1^{'}, \ldots, \eta_n^{'})$. Define the demand for college j as the fraction of students who would be assigned to college j, that is,

$$D_{n,j}(x^n, z_j^n, \epsilon^n, \eta^n; z_{-j}^n, p) = \frac{1}{n}\sum_{i=1}^{n} m_{ij}(x_i, \epsilon_i, \eta_i, z_{ij}; z_{i,-j}, p).$$

I can then pin down the $J \times 1$ vector of equilibrium cutoffs p_n via the following $J \times 1$ equilibrium conditions, for each college $j \in \{1, \ldots, J\}$,

$$D_{n,j}(x^n, z_j^n, \epsilon^n, \eta^n; z_{-j}^n, p_n) = s_j. \tag{1.3}$$

The equilibrium cutoffs $p_n(x^n, \epsilon^n, \eta^n, z^n, s)$ equates the supply and demand of the colleges and thus depend on all the characteristics of all the participants in the market. The endogeneity problem arises because students' taste shocks would influence their own choice set through equilibrium. This motives us to focus on large market where such correlation vanishes as n goes to infinity. Furthermore, the large market helps tackle the multiplicity of equilibria. There are generally multiple equilibria in small or finite markets but they converge to a unique equilibrium when the market size goes to large (Azevedo and Leshno, 2016).

1.3 Uniqueness and Convergence of Equilibrium

In this section, I discuss the properties of stable matchings in a limit continuum model. Three features make the limit approximation a convenient tool for identification and estimation. First, in the limit model, the equilibrium is deterministic as it depends on the distribution of the characteristics rather than the values of the characteristics of all the students and colleges. Second, a unique equilibrium exists in the limit model (Azevedo and Leshno, 2016). Third, the equilibria in the finite market converge to the equilibrium in the limit market (Azevedo and Leshno, 2016).

When $n \to \infty$, for any given p, the limiting demand function of college j can be written as

$$D_j(p) = E[m_{ij}(x_i, \epsilon_i, \eta_i, z_{ij}; z_{i,-j}, p)],$$

where the expectation is taken with respect to $(x_i, z_i, \epsilon_i, \eta_i)$. The vector of equilibrium cutoffs p^* is determined by the continuum analogue of Equations (1.3):

$$D_j(p) = s_j. \tag{1.4}$$

In the limiting economy, p^* only depends on the distribution of student's characteristics rather than the collection of all the individual students' characteristics.

1.3.1 Uniqueness

To establish the uniqueness of the limiting equilibrium, I impose assumptions on the joint distribution of unobserved error terms and on colleges' utility functions.

Assumption 1.

(i) (ϵ_i, η_i) are i.i.d. across i with CDF F, which is continuously differentiable.

(ii) $v^j(\cdot)$ is continuously differentiable in its last argument.

Assumption 1(i) requires independence across students. However, it is still quite flexible since it allow correlation between ϵ_i and η_i among colleges. The smoothness restrictions imposed in Assumption 1(i)(ii) guarantee that colleges have continuous preference.

Lemma 1. *(Azevedo and Leshno, 2016) Under Assumption 1, the limiting economy has a unique vector of equilibrium cutoff p^*.*

Proof. See Appendix. □

1.3.2 Convergence

Since the existence and uniqueness of a stable matching are well established in a continuum model, it is convenient to exploit the continuum economy as a tool to study large finite economies. The question is whether the stable matchings in large finite economies are well approximated by the stable matching in the continuum economy. Proposition 1 answers this question by showing that the set of equilibrium cutoffs in finite markets converge to the unique limit equilibrium cutoff in the Hausdorff distance as n goes to infinity.

Proposition 1. *(Azevedo and Leshno, 2016) Let $\mathcal{P}_n$ denote the set of equilibrium cutoffs of finite economy of size n. Under Assumptions 1 , given any s, as $n \to \infty$,*

$$\sup_{p_n \in \mathcal{P}_n} ||p_n(x^n, z^n, \epsilon^n, \eta^n) - p^*|| \xrightarrow{p} 0$$

Proof. See the Appendix. □

The basic idea of the proof is first to show that for any given, the demand function $D_n(p)$ in the finite economy converges in probability to $D(p)$ in the continuum economy by uniform law of large numbers. Then, under the continuity of the limit demand function $D(p)$, which is implied by Assumption 1, the equilibrium cutoffs that equating demand and supply in the finite economy must be close enough to that in the limit economy.

1.4 Nonparametric Identification

Given the convergence result in the previous section, for the rest of the paper, I analyze the identification and estimation of the limiting model. The model is a discrete choice model with unobserved and heterogeneous choice sets that depend on equilibrium. Although the identification of discrete choice models has been well established (McFadden, 1973, 1981; Manski, 1975, 1985; Cosslett, 1983; Horowitz, 1992; Ichimura, 1993; Klein and Spady, 1993; Lewbel, 2000; Berry and Haile, 2009; Chiappori and Komunjer, 2009; Briesch, Chintagunta, and Matzkin, 2010; Matzkin, 1991, 1992, 1993, 2007, 2012, 2018, among others), the identification of my model is challenging because matching markets are bilateral, in the sense that in order to be enrolled in a college, the student has to be admitted by this college; so the equilibrium outcome is determined by the two sides' preferences.

In this section, I discuss the nonparametric identification of the derivatives of the utility function as well as the joint distribution of the unobserved random terms. My identification analysis relies on exclusive regressors on both sides of the market, denoted by $(y_{i1}, \dots, y_{iJ})$ and $(w_{i1}, \dots, w_{iJ})$, respectively. y_{ij} only affects student i's preference over college j without influencing student i's choice set or altering her preferences over colleges other than j.[8] Similarly, w_{ij} is used to capture the likelihood that college j shows up in the choice set without changing student's preferences. Here, I consider a separable model with additive exclusive regressors. Specifically, for $j = 1, \dots, J$, the utility functions are given as

$$u_{ij} = u^j(x_i, z_{ij}) + y_{ij} + \epsilon_{ij}$$

$$v_{ji} = v^j(z_{ij}, x_i) + w_{ij} + \eta_{ji}.$$

For each j, utility functions of both sides are separable into an index of the form $\epsilon_{ij} + y_{ij}$

[8] A commonly used exclusive regressor y_{ij} in the literature (e.g., Agarwal, 2015) is the distance between student i and college j. The exclusive regressor on the other side, w_{ij}, for example, could be the decomposition of the admission committee, which will not affect student's preference since it is usually unobserved by students; or it could be the matching degree between worker's skill set and the firm's type.

and $\eta_{ji} + w_{ij}$, respectively. The coefficients of y_{ij} and w_{ij} are normalized to 1. Given that I do not impose any parametric assumption on the joint distribution of the unobserved terms, this is just a necessary scale normalization of utilities. I normalize the utility of not going to college to zero.[9] Separability is not necessary for the identification results; however, I focus on this specification to simplify the expressions and to illustrate the main idea. It's easy to extend this model to a more general nonseparable model based on Matzkin (2018) given that value of the outside option is observed, which is not needed in the current setting. Specifically, the utility functions and the joint CDF of the unobservables are also identified in the following nonseparable setting: for $j = 1$, the utility functions are given as

$$
\begin{aligned}
u_{i1} &= u^1(x_i, z_{i1}, \epsilon_{i1} + y_{i1}) \\
v_{1i} &= v^1(z_{i1}, x_i, \eta_{1i} + w_{i1});
\end{aligned}
$$

for $j = 2, \ldots, J$, the utility functions are given as

$$
\begin{aligned}
u_{ij} &= u^j(x_i, z_{ij}, y_{ij}, \epsilon_{ij}) \\
v_{ji} &= v^j(z_{ij}, x_i, w_{ij}, \eta_{ji}).
\end{aligned}
$$

1.4.1 Identifying the Derivative of the Utility Function

First consider identifying the derivatives of the utility functions. Assume that I observe the conditional probability of being matched to college j for all $j = 1, \ldots, J$. [10]

[9] My identification results in this subsection does not depend on the specific constant value assigned to the outside option because they rely on the derivatives of the conditional probabilities.

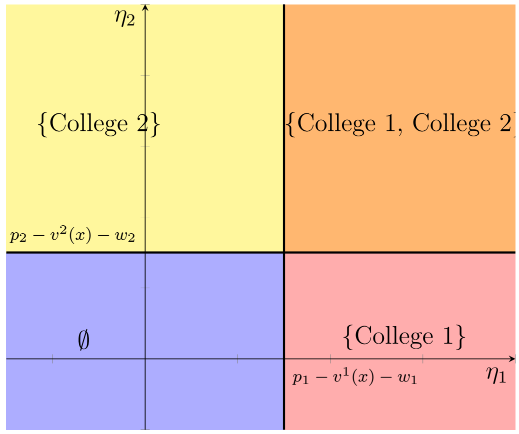
[10] The idea also applies if one only observes one college's conditional matched probability or the conditional probability of being unmatched. Given all the colleges' conditional probabilities and sufficient variations in the derivatives of the conditional probabilities, the model is overidentified. In this section, I use the minimal and necessary conditions for identification. I will revisit the overidentification problem when discussing semiparametric estimation of the model.

Overview and Intuition

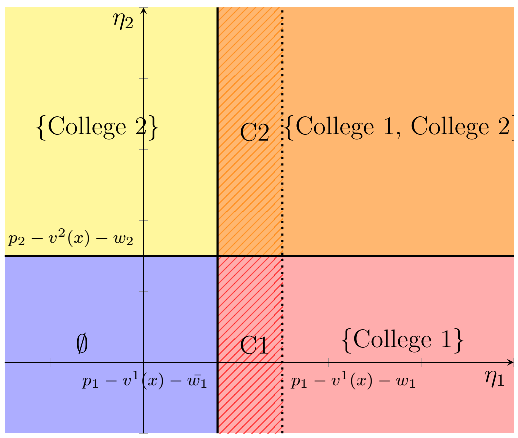
To illustrate the identification strategy, consider the special case of $J = 2$ colleges indexed by $j \in \{1, 2\}$. Here, for this simplified case, I illustrate how the exclusive regressors are used to facilitate the identification by tracing out the variation in the choice sets. Assume that the utility functions of college 1 and college 2 are $v_1 = v^1(x) + w_1 + \eta_1$ and $v_2 = v^2(x) + w_2 + \eta_2$, respectively. Assume that the (*unobserved*) cutoffs of college 1 and college 2 are p_1 and p_2, respectively. In Figure 1.1, I show the choice set of students in this game. Panel (A) shows that for any value of (x, w_1, w_2), student's choice set partitions R^2 into four parts depending on whether a student would choose from both colleges, only college 1, only college 2, or no college. A student would be admitted by college 1 if $\eta_1 > p_1 - v^1(x) - w_1$ and be admitted by college 2 if $\eta_2 > p_2 - v^2(x) - w_2$.

Panel (B) shows how distinct values of w_1 would lead to different choice sets, holding (x, w_2) fixed. Since w_1 determines the preference of college 1 over the student, the increase from w_1 to $\bar{w}_1$ makes college 1 more feasible. The areas C_1 and C_2 highlight the sets of students whose choice sets differ given these two different values of w_1. Since w_1 only enters v_1 and moves in the same way as η_1, the variation in w_1 will trace out the variation in η_1, without having to observe p_1. To see how the predicted conditional probabilities would change as the choice sets change, without loss of generality, assume that college 1 $\succ$ college 2 for all the students. Then the change in the conditional probability of being unmatched would depend on the mass that the density of (η_1, η_2) puts on the area C_1. Similarly, the change in the conditional probability of being matched with college 2 would depend on the mass that the density of (η_1, η_2) puts on the area C_2. One can draw a similar graph for w_2, the change of which would move the partition along the vertical axis.

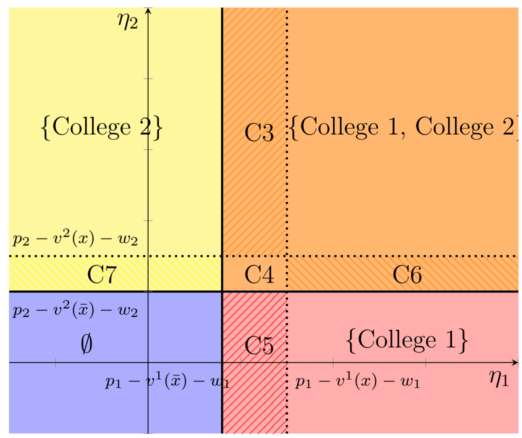
Panel (C) depicts the change in the choice sets induced by shifting x. Since x enters the utilities of both colleges, v_1 and v_2, the change in it moves the partition along both axes. Different from w_1 and w_2, the way that x moves the partition depends on the shape of the functions, $v^1(\cdot)$ and $v^2(\cdot)$. Areas C_3 to C_7 show the sets of students who face different choice



(a) Choice set given (x, w_1, w_2)



(b) Change in choice set induced by w_1



(c) Change in choice set induced by x

Figure 1.1: Choice Set Partitions

sets under the two distinct values of x. Similar to panel (B), one can connect the change in the predicted conditional probabilities with the change in the partitions by considering the preferences of students and the mass placed on each set.

The identification results, as will be formally shown below, illustrate the trade off between the changes caused by the variables. In this simplified case, for $j = 0$, 1, 2, the conditional probability of choosing the option j satisfy the following equation

$$\frac{\partial P(m_j|x, w_1, w_2)}{\partial x} = \frac{\partial v^1}{\partial x}\frac{\partial P(m_j|x, w_1, w_2)}{\partial w_1} + \frac{\partial v^2}{\partial x}\frac{\partial P(m_j|x, w_1, w_2)}{\partial w_2}. \tag{1.5}$$

This expression reflects the chain rule: the effect of x on the conditional probability of being matched with college j, $P(m_j|x, w_1, w_2)$, is realized through its effects on latent utilities v_j, captured by $\frac{\partial v^1}{\partial x}$ and $\frac{\partial v^2}{\partial x}$. $\frac{\partial P(m_j|x,w_1,w_2)}{\partial w_1}$ and $\frac{\partial P(m_j|x,w_1,w_2)}{\partial w_2}$ capture the effect of latent utilities on the conditional probability of being matched with college j. There are two unknown elements in Eq. (1.5). If (η_1, η_2) has enough variation such that fixing the value of x, distinct values of (w_1, w_2) produce two such equations, one can solve the problem.

Formal Identification Results

Assumption 2.

(i) (ϵ_i, η_i) *are distributed independently of* $(x_i, z_{i1}, \ldots, z_{iJ}, y_{i1}, \ldots, y_{iJ}, w_{i1}, \ldots, w_{iJ})$*;*

(ii) for all j*,* u^j *and* v^j *functions are continuously differentiable.*

Assumption 2(i) imposes an independence assumption. The insights presented below still apply when extending this model using control method approach or other methods that deal with the endogeneity problem. Moreover, in the generalized version of my model where the unobserved terms enter the utility functions nonadditively, the assumption is much weaker than it would be if they enter additively. The differentiability of the utility functions, u^j and v^j, and the CDF of (ϵ_i, η_i), F, in Assumption2 (ii) and Assumption 1(ii) in Section 3, guarantees that the conditional probability is differentiable.

A student ends up with college j if and only if (i) the student prefers college j over the outside option and is admitted by college j, and (ii) for each j', either the student is not admitted by college j', or the student prefers college j over college j', or both. Specifically, for any given value $(x, z_1, \ldots, z_J, y_1, \ldots, y_J, w_1, \ldots, w_J, \epsilon_1, \ldots, \epsilon_J, \eta_1, \ldots, \eta_J)$ in the support of $(x_i, z_{i1}, \ldots, z_{iJ}, y_{i1}, \ldots, y_{iJ}, w_{i1}, \ldots, w_{iJ}, \epsilon_{i1}, \ldots, \epsilon_{iJ}, \eta_{1i}, \ldots, \eta_{Ji})$, the indicator of student i

being matched with college j can be written as

$$
\begin{aligned}
& m_j(x, z_1, \ldots, z_J, y_1, \ldots, y_J, w_1, \ldots, w_J, \epsilon_1, \ldots, \epsilon_J, \eta_1, \ldots, \eta_J; p) \\
= & \begin{cases}
1 & \text{if } \underbrace{\epsilon_j > -u^j(x, z_j) - y_j}_{\text{college } j \succ \text{The outside option}} \quad or \ \underbrace{\eta_j > p_j - v^j(z_j, x) - w_j}_{\text{Admitted by college } j} \\
& \text{and for all } j' \neq j, \underbrace{\epsilon_j - \epsilon_{j'} > -u^j(x, z_j) - y_j + u^{j'}(x, z_{j'}) + y_{j'}}_{\text{college } j \succ \text{college } j'} \\
& \text{or } \underbrace{\eta_{j'} < p_{j'} - v^{j'}(z_{j'}, x) - w_{j'}}_{\text{Not admitted by college } j'} \\
0 & otherwise
\end{cases}
\end{aligned}
$$

Integrating m_j with respect to all the unobserved terms, one gets the conditional probability of the student being unmatched:

$$
\begin{aligned}
& \sigma_j(x, z_1, \ldots, z_J, y_1, \ldots, y_J, w_1, \ldots, w_J; p) \\
= & \int m_j(x, z_1, \ldots, z_J, y_1, \ldots, y_J, w_1, \ldots, w_J, \epsilon_1, \ldots, \epsilon_J, \eta_1, \ldots, \eta_J; p) \\
& dF(\epsilon_1, \ldots, \epsilon_J, \eta_1, \ldots, \eta_J) \\
=: & \Lambda_j(-u^1(x, z_1) - y_1, \ldots, -u^J(x, z_J) - y_J, \\
& p_1 - v^1(z_1, x) - w_1, \ldots, p_J - v^J(z_J, x) - w_J).
\end{aligned}
$$

Here, Λ_j is some unknown function. The integration over the unobservables does not change the additive form of the exclusive regressors in the utility functions.

I now extend the insights in the previous subsection to the general case of J colleges. Let $(x, z_1, \ldots, z_J)$ be a given value in the support of $(x_i, z_{i1}, \ldots, z_{iJ})$. Consider 2 different values for the exclusive regressors $(y_{i1}, \ldots, y_{iJ}, w_{i1}, \ldots, w_{iJ})$, $(y_{i1}^{(1)}, \ldots, y_{iJ}^{(1)}, w_{i1}^{(1)}, \ldots, w_{iJ}^{(1)})$ and $(y_{i1}^{(2)}, \ldots, y_{iJ}^{(2)}, w_{i1}^{(2)}, \ldots, w_{iJ}^{(2)})$. For $k = 1, 2$, define

$$
\sigma_j^{(k)} = P(m_{ij} = 1 | x, z_1, \ldots, z_j, y_1^{(k)}, \ldots, y_J^{(k)}, w_1^{(k)}, \ldots, w_J^{(k)}; p).
$$

Let $x_{(l)}$ denote a coordinate of x. As will be shown in the proof, I obtain the following system of equations, which follows similar idea as in the simplified example in Eq. (1.5),

$$\underbrace{\begin{bmatrix} \frac{\partial \sigma_1^{(1)}}{\partial x_{(l)}} \\ \vdots \\ \frac{\partial \sigma_J^{(1)}}{\partial x_{(l)}} \\ \frac{\partial \sigma_1^{(2)}}{\partial x_{(l)}} \\ \vdots \\ \frac{\partial \sigma_J^{(2)}}{\partial x_{(l)}} \end{bmatrix}}_{T_{x_{(l)}}} = \underbrace{\begin{bmatrix} \frac{\partial \sigma_1^{(1)}}{\partial y_1} & \cdots & \frac{\partial \sigma_1^{(1)}}{\partial y_J} & \frac{\partial \sigma_1^{(1)}}{\partial w_1} & \cdots & \frac{\partial \sigma_1^{(1)}}{\partial w_J} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \sigma_J^{(1)}}{\partial y_1} & \cdots & \frac{\partial \sigma_J^{(1)}}{\partial y_J} & \frac{\partial \sigma_J^{(1)}}{\partial w_1} & \cdots & \frac{\partial \sigma_J^{(1)}}{\partial w_J} \\ \frac{\partial \sigma_1^{(2)}}{\partial y_1} & \cdots & \frac{\partial \sigma_1^{(2)}}{\partial y_J} & \frac{\partial \sigma_1^{(2)}}{\partial w_1} & \cdots & \frac{\partial \sigma_1^{(2)}}{\partial w_J} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \sigma_J^{(2)}}{\partial y_1} & \cdots & \frac{\partial \sigma_J^{(2)}}{\partial y_J} & \frac{\partial \sigma_J^{(2)}}{\partial w_1} & \cdots & \frac{\partial \sigma_J^{(2)}}{\partial w_J} \end{bmatrix}}_{\Pi} \times \underbrace{\begin{bmatrix} u_{x_{(l)}}^1 \\ \vdots \\ u_{x_{(l)}}^J \\ v_{x_{(l)}}^1 \\ \vdots \\ v_{x_{(l)}}^J \end{bmatrix}}_{\beta(x_{(l)})}, \tag{1.6}$$

where $u_{x_{(l)}}^j = \frac{\partial u^j}{\partial x_{(l)}}$ and $v_{x_{(l)}}^j = \frac{\partial v^j}{\partial x_{(l)}}$. The matrix $\Pi(x, z, y^{(1)}, y^{(2)}, w^{(1)}, w^{(2)}; p)$ is formed by fixing the values of the student's characteristics x, fixing the values of the characteristics of all colleges, $z_1, \ldots, z_J$. The matrix depends on the derivatives with respect to the exclusive regressors $(y_1, \ldots, y_J, w_1, \ldots, w_J)$ of the probability of being matched to college $j = 1, \ldots, J$. These derivatives are calculated at 2 different values of $(y_1, \ldots, y_J, w_1, \ldots, w_J)$.

The key identification condition is shown as follows:

Condition 1. *Given $(x, z_1, \ldots, z_J)$ there exists 2 values, $(y_1^{(1)}, \ldots, y_J^{(1)}, w_1^{(1)}, \ldots, w_J^{(1)})$ and $(y_1^{(2)}, \ldots, y_J^{(2)}, w_1^{(2)}, \ldots, w_J^{(2)})$, in the support of $(y_{i1}, \ldots, y_{iJ}, w_{i1}, \ldots, w_{iJ})$ conditional on $(x_i, z_{i1}, \ldots, z_{iJ})= (x, z_1, \ldots, z_J)$ such that*

$$\Pi(x, z; y^{(1)}, y^{(2)}, w^{(1)}, w_1^{(2)}; p) \text{ has rank } 2J.$$

Denote $\Pi_{x_{(l)}}^j$ to be the matrix formed by replacing the $j-th$ column of matrix Π by the vector $T_{x_{(l)}}$. Let $z_{j(l)}$ denote a coordinate of z_j. The next proposition, which exploits similar ideas as in Matzkin (2018), establishes a constructive, pointwise identification at a given $(x, z_1, \ldots, z_J)$ vector, of the derivatives of u^j and v^j functions for each $j = 1, \ldots, J$.

Proposition 2. *Suppose that Assumptions 1, 2, and Condition 1 are satisfied. Then, for* *all* $j = 1, \ldots, J$, *for any* l,

$$\frac{\partial u^j(x, z_j)}{\partial x_{(l)}} = \frac{|\Pi|}{|\Pi^j_{x_{(l)}}|}, \quad \frac{\partial v^j(z_j, x)}{\partial x_{(l)}} = \frac{|\Pi|}{|\Pi^{j+J}_{x_{(l)}}|}.$$

Moreover, for any $j_1, j_2 = 1, \ldots, J$, *given that* $\frac{\partial \sigma_{j_1}}{\partial y_j}\frac{\partial \sigma_{j_1}}{\partial w_j} - \frac{\partial \sigma_{j_2}}{\partial y_j}\frac{\partial \sigma_{j_2}}{\partial w_j} \neq 0$,

$$\frac{\partial u^j(x, z_j)}{\partial z_j} = \Big(\frac{\partial \sigma_{j_1}}{\partial z_j}\frac{\partial \sigma_{j_2}}{\partial w_j} - \frac{\partial \sigma_{j_2}}{\partial z_j}\frac{\partial \sigma_{j_1}}{\partial w_j}\Big)\Big/\Big(\frac{\partial \sigma_{j_1}}{\partial y_j}\frac{\partial \sigma_{j_2}}{\partial w_j} - \frac{\partial \sigma_{j_2}}{\partial y_j}\frac{\partial \sigma_{j_1}}{\partial w_j}\Big),$$

$$\frac{\partial v^j(z_j, x)}{\partial z_j} = \Big(\frac{\partial \sigma_{j_1}}{\partial z_j}\frac{\partial \sigma_{j_2}}{\partial y_j} - \frac{\partial \sigma_{j_2}}{\partial z_j}\frac{\partial \sigma_{j_1}}{\partial y_j}\Big)\Big/\Big(\frac{\partial \sigma_{j_1}}{\partial y_j}\frac{\partial \sigma_{j_2}}{\partial w_j} - \frac{\partial \sigma_{j_2}}{\partial y_j}\frac{\partial \sigma_{j_1}}{\partial w_j}\Big).$$

Proof. See the Appendix. □

Proposition 2 shows how to identify the derivatives of the utility functions of the student u^j and of the college v^j for each j from the probabilities of being matched to the colleges. Assuming $(x, z_1, \ldots, z_J)$ have full support, under a location normalization, for each j, the values of u^j and v^j are identified.

1.4.2 Identifying the Joint Distribution of Unobservables

In this subsection, given that the values of the utility functions u^j and v^j are known for each j, I analyze how to recover the joint distribution of unobservables. I focus on the conditional probability of being unmatched . In a classical (one-sided) discrete choice model, the conditional probability of being unmatched is

$$\begin{aligned} &P(m_0|x, z, y) \\ =&P(\epsilon_1 \leq -u^1(x, z_1) - y_1, \cdots, \epsilon_J \leq -u^J(x, z_J) - y_J) \\ =&F(-u^1(x, z_1) - y_1, \cdots, -u^J(x, z_J) - y_J). \end{aligned}$$

So once the utility functions are identified, it is straightforward to identify the CDF of the unobserved terms. However, it is not the case in a two-sided matching model. A student not going to any college may be because she does not want to attend any college, or because she is not qualified for any college that she wants to attend. Formally, note that the indicator of being unmatched

$$
\begin{aligned}
&m_0(x, z, y, w, \epsilon, \eta; p)\\
&= \begin{cases} 1 & \text{if for all } j \geq 1,\ \epsilon_j \leq -u^j(x, z_j) - y_j \textit{ or } \eta_j \leq p_j - v^j(z_j, x) - w_j \\ 0 & \textit{otherwise} \end{cases}.
\end{aligned}
$$

Therefore, the conditional probability of being unmatched can be written as

$$
\begin{aligned}
&P(m_0 = 1 | x, z, y, w; p)\\
=&P\Big(\cap_{j=1}^{J} \{(\epsilon_j \leq -u^j(x, z_j) - y_j) \cup (\eta_j \leq p_j - v^j(z_j, x) - w_j)\}\Big)\\
=:&\Lambda_0(-u^1(x, z_1) - y_1, \ldots, -u^J(x, z_J) - y_J, p_1 - v^1(z_1, x) - w_1, , \ldots, p_J - v^J(z_J, x) - w_J).
\end{aligned}
\tag{1.7}
$$

Once u^j and v^j are identified, since the LHS in (1.7) can be recovered from the data, $\Lambda_0()$ in (1.7) is also identified. The key is to find the relationship between $\Lambda_0()$ and the joint CDF F.

Overview and Intuition

To illustrate the main idea of identifying the joint CDF of the unobserved terms, consider a special case with $J = 1$. For this example, assume the utility of attending college is $u = u(x) + y + \epsilon$, the utility of not attending college is 0. The utility of the college over the student is given by $v = v(x) + w + \eta$. In Figure 1.2, I show matching outcomes in this game. Given any value of (x, y, w), student's matching outcomes partition R^2 into two parts, college or no college. I focus on the conditional probability of no college to identify the CDF, which corresponds to the hatch area on the graph. This area can be decomposed into three parts, R_1, R_2, and R_3. It's easy to see that

$$\begin{aligned}P(\text{No College}) =&P(\epsilon < -u(x) - y \text{ or } \eta < p - v(x) - w)\\ =&P(R_1 \cup R_2) + P(R_3 \cup R_2) - P(R_2).\end{aligned}$$

The last term in the above expression is indeed the joint CDF of the unobserved terms we want to recover since $P(R_2) = P(\epsilon < -u(x)-y, \eta < p-v(x)-w) = F(-u(x)-y, p-v(x)-w)$. Therefore, as long as we can recover $P(R_1 \cup R_2)$ and $P(R_3 \cup R_2)$, since $P(\text{No College})$ on the LHS can be estimated from the data, the joint CDF of the unobserved terms is identified. Note that $P(R_1 \cup R_2) = F_\epsilon(-u(x) - y)$ is actually the marginal CDF of ϵ. It can be recovered by considering the subset of students whose w (e.g., SAT scores) is high enough so that these students would be qualified for college regardless of the realization of the unobserved ability η. In this way, the effects of college's preferences is "shut down" and $F_\epsilon(-u(x) - y) = P(\epsilon < -u(x) - y) = P(\text{No College}|w \text{ is high})$ is identified. Similarly, F_η can be identified and as a result, the joint CDF, F is identified.

Formal Identification Results

Consider (1.7) for the general case. It's easy to see that the joint distribution of unobserved heterogeneity in student's preferences, the CDF of $(\epsilon_{i1}, \ldots, \epsilon_{iJ})$ can be recovered by considering the conditional probability of being unmatched of the sub-sample with large enough values of $(p_1 - v^1(z_{i1}, x_i) - w_{i1}, \ldots, p_J - v^J(z_{iJ}, x_i) - w_{iJ})$, and similar for the other side. Since I does not restrict the unobserved random terms $(\epsilon_{i1}, \ldots, \epsilon_{iJ}, \eta_{1i}, \ldots, \eta_{Ji})$ to be independent, to fully identify the entire distribution of all unobserved terms, in my formal proof, I show that $\Lambda_0()$ can be expressed as the sum of F, the joint CDF of $(\epsilon_{i1}, \ldots, \epsilon_{iJ}, \eta_{1i}, \ldots, \eta_{Ji})$, and the joint CDFs of several subsets of $(\epsilon_{i1}, \ldots, \epsilon_{iJ}, \eta_{1i}, \ldots, \eta_{Ji})$. The intuition behind this, as illustrated in Figure 1.2, is that a student not going to any college can be attributed to

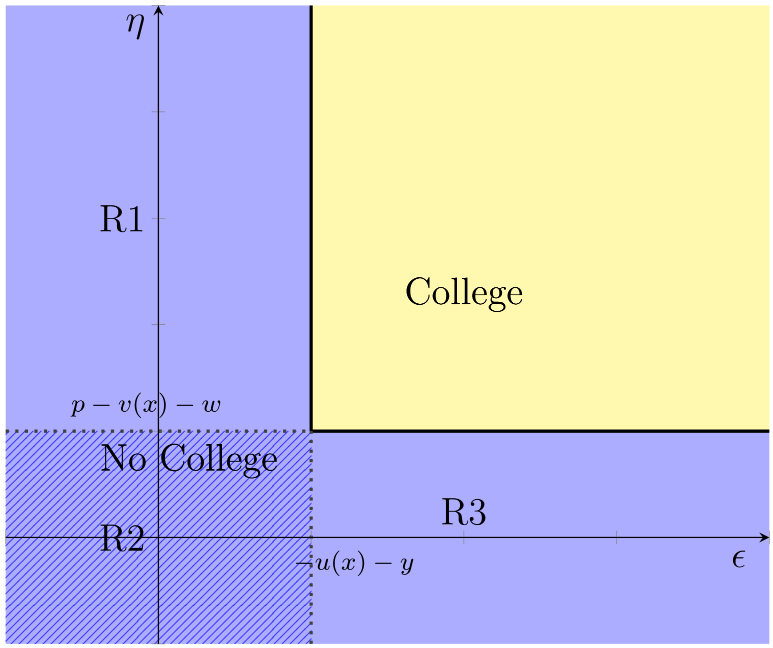


Figure 1.2: Decomposition of Matching Probability

several scenarios. One is that the student does not prefer any college over the outside option *and* is not qualified for any college. The probability of this scenario corresponds to the joint CDF of the unobserved terms we want to recover. The other scenarios include, for example, that the student is not interested in the colleges that she is admitted by and is rejected by the colleges that she wants to attend. For all these scenarios, one can express the conditional probability by the joint CDF of a subset of the unobserved terms. So F can be identified by subtracting the conditional probabilities of various scenarios of being unmatched, recovered by "shutting down" the effects the corresponding subsets of the unobservables, from the conditional probability of being unmatched. The following proposition states the identification result.

Proposition 3. *Suppose for each $j = 1, \ldots, J$, u^j and v^j are identified. Then F, the joint CDF of $(\epsilon_{i1}, \ldots, \epsilon_{iJ}, \eta_{1i}, \ldots, \eta_{Ji})$ is identified.*

Proof. See the Appendix. □

1.5 Estimation

1.5.1 Nonparametric Estimation

The constructive identification results in the previous section facilitate the estimation for the unknown values of coefficients in (1.6). Let ω denote the vector of all the observed covariates, i.e., $\omega = (x, z_1, \ldots, z_J, y_1, \ldots, y_J, w_1, \ldots, w_J)$, and k_ω be the dimension of ω. Each of the elements in the matrices in (1.6) can be estimated by replacing the conditional probability $\sigma_j(\omega)$ with a nonparametric estimator, $\hat{\sigma}_j(\omega)$, for $\sigma_j(\omega)$. Denote the estimators for Π and $T_{x_{(l)}}$ by $\widehat{\Pi}$ and $\widehat{T}_{x_{(l)}}$, respectively. Then the estimator for the vector of derivatives $\beta(x_{(l)})$ is defined as

$$\hat{\beta}(x_{(l)}) = \widehat{\Pi}^{-1}\widehat{T}_{x_{(l)}}. \tag{1.8}$$

In this section, we develop asymptotic properties for the estimator presented in (1.8) when the conditional probability $\sigma_j(\omega)$ is estimated using Kernel methods. The kernel estimator is

$$\hat{\sigma}_j(\omega) = \frac{\sum_{i=1}^{n} m_{ij} K(\frac{\omega_i - \omega}{h_n})}{\sum_{i=1}^{n} K(\frac{\omega_i - \omega}{h_n})}, \tag{1.9}$$

where K is a kernel function and h_n is a bandwidth.

I will make the following assumptions:

Assumption 3.

(i) *The density f_ω has compact support and is continuously differentiable of order d.*

(ii) *The kernel function K is differentiable of order Δ, the derivatives of order Δ are Lipschitz, where $\Delta \geq 1$. K vanishes outside a compact set, integrates to 1, and is of order s, where $s + 1 \leq d$.*

(iii) $0 < f_\omega < \infty$.

(iv) The sequence of bandwidths, h_n, is such that $h_n \to 0$, $nh_n^{k_\omega+2} \to \infty$, $\sqrt{nh_n^{k_\omega+2}}h_n^s \to 0$, $[(nh_n^{2k_\omega+2})/\ln(n)] \to \infty$, and $\sqrt{nh_n^{k_\omega+2}}(\sqrt{\ln(n)/(nh_n^{2k_\omega+2})} + h_n^s)^2 \to 0$.

Assumption 3(i) requires the pdf of ω to be sufficiently smooth. Assumption 3(ii) requires the Kernel function to be of high order and together with Assumption 3 (i), guarantees that the bias of the estimator vanishes in the limit. Assumption 3(iii) assumes that the pdf of ω is bounded and bounded away from zero. Assumption 3(iv) restricts the rate at which the bandwidth h_n goes to zero.

Define

$$\tilde{K}(\omega) = (K(\omega), \frac{\partial K(\omega)}{\partial y_1}, \ldots, \frac{\partial K(\omega)}{\partial y_J}, \frac{\partial K(\omega)}{\partial w_1}, \ldots, \frac{\partial K(\omega)}{\partial w_J}, \frac{\partial K(\omega)}{\partial x_{(l)}})',$$

$$\Sigma(\omega) = \begin{bmatrix} 1 & \sigma_1(\omega) & \cdots & \cdots & \sigma_J(\omega) \\ \sigma_1(\omega) & \sigma_1(\omega) & 0 & \cdots & 0 \\ \vdots & 0 & \sigma_2(\omega) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_J(\omega) & 0 & 0 & \cdots & \sigma_J(\omega) \end{bmatrix}.$$

The $2J \times (2J+2)(J+1)$ matrix Γ is defined by (1.40)-(1.46) in the proof of Proposition 4 in the Appendix.

Let

$$V = \Gamma[\Sigma(\omega) \otimes \int \tilde{K}(\tilde{\omega})\tilde{K}(\tilde{\omega})' d\tilde{\omega}]\Gamma'. \tag{1.10}$$

and

$$\hat{V} = \hat{\pi}(\omega)'[\hat{\Sigma}(\omega) \otimes \int \tilde{K}(\tilde{\omega})\tilde{K}(\tilde{\omega})' d\tilde{\omega}]\hat{\pi}(\omega) \tag{1.11}$$

be an estimator for V, obtained by substituting the density and the conditional probability of being matched by their kernel estimators, respectively, in the definition of Γ, and substituting σ_j by $\hat{\sigma}_j$, defined in (1.9), in the definition of Σ. The following proposition establishes the asymptotic distribution of the nonparametric estimator $\hat{\beta}(x_{(l)})$ defined in (1.8).

Proposition 4. *Suppose Assumptions 1-3 and Condition 1 hold. Then,*

$$\sqrt{nh_n^{k_\omega+2}}(\hat{\beta}(x_{(l)}) - \beta(x_{(l)})) \xrightarrow{d} N(0, V)$$

and $\hat{V}$ is a consistent estimator for V.

1.5.2 Semiparametric Estimation

Although nonparametric estimation imposes less restrictive assumptions on the functional form of the utilities, the nonparametric estimators inevitably converge slowly. The convergence rate of the estimators decreases with the number of variables involved, so called "curse of dimensionality." This issue is particularly challenging in the context of matching as the number of colleges or firms is usually large. In this section, I consider a semiparametric model, where the utility functions are specified parametrically, and the distribution of the unobserved heterogeneity is still not parametrically specified. Based on the nonparametric identification results in Section 4 and the idea in Powell, Stock, and Stoker (1989), I derive moment conditions where the elements other than the parameters of interest can be estimated using average derivative estimators. This allows us to propose a GMM estimator that is root-n consistency and asymptotically normal.

The model considered in this section is, for each $j = 1, \dots, J$,

$$\begin{aligned} u_{ij} &= \beta' x_i + \gamma' z_{ij} + y_{ij} + \epsilon_{ij} \\ v_{ji} &= \alpha' x_i + \rho' z_{ij} + w_{ij} + \eta_{ji}. \end{aligned}$$

The parameter of interest is $\theta = (\beta, \alpha, \gamma, \rho)$. Here I still do not make any parametric assumptions on the joint distribution of $(\epsilon_{i1}, \dots, \epsilon_{iJ}, \eta_{i1}, \dots, \eta_{iJ})$. Moreover, I assume that the cutoffs are unobserved. The utility of the outside option is normalized to 0.

Denote the vector of all the observed covariates by ω_i, i.e., $\omega_i = (x_i, z_i, y_i, w_i)$, with $z_i =$

$(z_{i1}, \ldots, z_{iJ})$, $y_i = (y_{i1}, \ldots, y_{iJ})$ and $w_i = (w_{i1}, \ldots, w_{iJ})$. Let $\sigma_j(\omega) = P(m_{ij} = 1|\omega = \omega_i)$. My nonparametric identification results in section 4 imply the following moment conditions: for each $j = 0, 1, \ldots, J$ and $j' = 1, \ldots, J$,

$$\frac{\partial \sigma_j(\omega)}{\partial x} = \beta \sum_{j'=1}^{J} \frac{\partial \sigma_j(\omega)}{\partial y_{j'}} + \alpha \sum_{j'=1}^{J} \frac{\partial \sigma_j(\omega)}{\partial w_{j'}} \tag{1.12}$$

$$\frac{\partial \sigma_j(\omega)}{\partial z_j} = \gamma \frac{\partial \sigma_j(\omega)}{\partial y_j} + \rho \frac{\partial \sigma_j(\omega)}{\partial w_j}. \tag{1.13}$$

While $\beta' x_i$ is constant in all the utility functions of student i over colleges, β is still identified from the comparison between colleges and the outside option. Because the conditional probabilities $\sigma_0, \sigma_1, \ldots, \sigma_J$ sum up to one, I use only the moment conditions for $j = 1, \ldots, J$.

In the following, I maintain the same assumptions as in Powell, Stock, and Stoker (1989).

Assumption 4.

(i) The support Ω of ω is a convex, possibly unbounded, subset of $\mathbb{R}^{k_\omega}$ with a nonempty interior Ω_0.

(ii) The density function f is continuous in the components of ω for all $\omega \in \mathbb{R}^{d_\omega}$, so that $f(\omega) = 0$ for all $\omega \in \partial\Omega$, where $\partial\Omega$ denotes the boundary of Ω. Moreover, f is continuously differentiable in the components of ω for all $\omega \in \mathbb{R}^{d_\omega}$.

(iii) For each j, the components of the vector $\frac{\partial \sigma_j(\omega)}{\partial \omega}$ and of the matrix $[\frac{\partial f(\omega)}{\partial \omega}](m_j, \omega')$ have finite second moments. Besides, $\frac{\partial f(\omega)}{\partial \omega}$ and $\frac{\partial [\sigma_j(\omega) f(\omega)]}{\partial \omega}$ satisfy the following Lipschitz conditions: for some $r(\omega)$,and

$$E[(1 + |m_j| + \|\omega\|)(r(\omega))]^2 < \infty,$$
$$\|\frac{\partial f(\omega + \zeta)}{\partial \omega} - \frac{\partial f(\omega)}{\partial \omega}\| < r(\omega)\|\zeta\|,$$

and

$$\|\frac{\partial [f(\omega + \zeta)\sigma_j(\omega + \zeta)]}{\partial \omega} - \frac{\partial [f(\omega)\sigma_j(\omega)]}{\partial \omega}\| < r(\omega)\|\zeta\|.$$

Assumption 4(i) implies that ω is continuously distributed and that no component of ω is functionally determined by other components.[11] Assumption 4(ii) gives a key boundary condition, which permits unbounded covariates and gives the smoothness conditions on the density function. Assumption 4(iii) requires the existence of various moments and imposes smoothness requirements on $f(\omega)$ and $\sigma_j(\omega)$.

The following lemma derives the moment conditions that I use to construct the GMM estimator for the parameter of interest.

Lemma 2. *Given Assumptions 2(ii) and 4, the conditions (1.12) and (1.13) imply the following conditions: for each j,*

$$E[m_j \frac{\partial f(\omega)}{\partial x}] = \beta \sum_{j=1}^{J} E[m_j \frac{\partial f(\omega)}{\partial y_{j'}}] + \alpha \sum_{j=1}^{J} E[m_j \frac{\partial f(\omega)}{\partial w_{j'}}] \tag{1.14}$$

$$E[m_j \frac{\partial f(\omega)}{\partial z_j}] = \gamma E[m_j \frac{\partial f(\omega)}{\partial y_j}] + \rho E[m_j \frac{\partial f(\omega)}{\partial w_j}]. \tag{1.15}$$

Proof. See the Appendix. □

Let the $k_g = J(k_x + k_{z_j})$ moment conditions

$$g(\theta) = B - A\theta \tag{1.16}$$

[11] Continuous random variables guarantee that the density is well-defined when doing integration by parts. When some covariates are discrete, the coefficients of these covariates can be estimated using the estimator developed by Horowitz and Härdle (1996). Alternatively, we can use full mean method by Newey (1994), which provides an estimator that has a bit slower convergence rate than the one discussed in this paper.

where

$$A = \begin{pmatrix} \left(\sum_{j=1}^{J} E[m_1 \frac{\partial f(\omega)}{\partial y_j}] \quad \sum_{j=1}^{J} E[m_1 \frac{\partial f(\omega)}{\partial w_j}]\right) \otimes I_{k_x} & 0_{k_x \times 2k_{z_1}} \\ 0_{k_{z_1} \times 2k_x} & \left(E[m_1 \frac{\partial f(\omega)}{\partial y_1}] \quad E[m_1 \frac{\partial f(\omega)}{\partial w_1}]\right) \otimes I_{k_{z_1}} \\ \vdots & \vdots \\ \left(\sum_{j=1}^{J} E[m_J \frac{\partial f(\omega)}{\partial y_j}] \quad \sum_{j=1}^{J} E[m_J \frac{\partial f(\omega)}{\partial w_j}]\right) \otimes I_{k_x} & 0_{k_x \times 2k_{z_J}} \\ 0_{k_{z_J} \times 2k_x} & \left(E[m_J \frac{\partial f(\omega)}{\partial y_J}] \quad E[m_J \frac{\partial f(\omega)}{\partial w_J}]\right) \otimes I_{k_{z_J}} \end{pmatrix}, \tag{1.17}$$

$$B = \begin{pmatrix} E[m_1 \frac{\partial f(\omega)}{\partial x}] \\ E[m_1 \frac{\partial f(\omega)}{\partial z_1}] \\ \vdots \\ E[m_J \frac{\partial f(\omega)}{\partial x}] \\ E[m_J \frac{\partial f(\omega)}{\partial z_J}] \end{pmatrix}, \tag{1.18}$$

and k_x and k_{z_j} denote the dimension of x and z_j, respectively.

The expectation terms in (1.17) and (1.18) can be estimated using the average derivative estimator:

$$\hat{E}[m_j \frac{\partial f(\omega)}{\partial \omega}] = \frac{1}{n} \sum_{i=1}^{n} m_{ij} \frac{\partial \hat{f}(\omega_i)}{\partial \omega_i}, \tag{1.19}$$

where $\frac{\partial \hat{f}(\omega)}{\partial \omega}$ is the kernel density estimator

$$\frac{\partial \hat{f}(\omega)}{\partial \omega} = \frac{1}{n-1} \frac{1}{h_n^{k_\omega+1}} \sum_{l \neq i} K'(\frac{\omega_l - \omega}{h_n}). \tag{1.20}$$

Substituting each expectation terms $E[m_j \frac{\partial f(\omega)}{\partial \omega}]$ with its estimator $\hat{E}[m_j \frac{\partial f(\omega)}{\partial \omega}]$ in (1.17) and (1.18), respectively, one obtains $\hat{A}_n$ and $\hat{B}_n$. The sample analogue of the moment conditions are

$$\hat{g}_n(\theta) = \hat{B}_n - \hat{A}_n \theta. \tag{1.21}$$

The GMM estimator for θ, denoted by $\hat{\theta}$, is defined as the minimizer of the GMM criterion

$Q_n(\theta) = -\hat{g}_n(\theta)'\hat{W}\hat{g}_n(\theta)$, where $\hat{W}$ is some positive semi-definite weight matrix and $\hat{W} \xrightarrow{p} W$. Specifically, $\hat{\theta}$ can be written as

$$\hat{\theta} = (\hat{A}\hat{W}\hat{A})^{-1}(\hat{A}\hat{W}\hat{B}). \tag{1.22}$$

Assumption 5.

(i) The kernel function K is symmetrical about the origin, bounded, differentiable, and of order s. $s = \frac{k_\omega+2}{2}$ if k_ω is even and $s = \frac{k_\omega+3}{2}$ if k_ω is odd, where k_ω denote the dimension of the covariates. In addition, $\int K(t)dt = 1$.

(ii) The bandwidth h_n satisfies $nh_n^{2s} \to 0$ and $nh_n^{k_\omega+2} \to \infty$ as $n \to \infty$.

Assumption 6.

(i) $A_j'WA_j$ is non-singular.

(ii) $E[\|\frac{\partial f(\omega)}{\omega}\|^2]$ is finite, where $\| \, \|$ refers to Euclidean norm.

Assumption 5(i) requires K to be a higher-order kernel to assure the bias of the average derivative estimators has size $o(n^{-1/2})$. Assumption 5(ii) imposes requirements on the convergence rate of the bandwidth h_n, which is relatively faster than the optimal convergence rate used in density estimation. Both assumptions are needed to prevent the asymptotic distribution of $\sqrt{n}(\hat{E}[m_{ij}\frac{\partial f(\omega)}{\partial \omega}] - E[m_{ij}\frac{\partial f(\omega)}{\partial \omega}])$ from having a nonzero mean. Assumption 6 is regularity conditions for GMM estimator.

The following proposition establishes the asymptotic properties of my GMM estimator.

Proposition 5. *Suppose Assumptions 1,2,4, 5, 6, and Condition 1 hold, then the GMM estimator, defined by (1.22), is consistent for θ. Furthermore, we have*

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \Sigma_\theta), \tag{1.23}$$

where

$$\Sigma_\theta = (A'WA)^{-1}A'WT\Sigma_\delta T^{'}WA(A'WA)^{-1}, \tag{1.24}$$

$$T = \begin{pmatrix} T_1 & 0 & \cdots & 0 \\ 0 & T_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & T_J \end{pmatrix}, \quad T_j = \begin{pmatrix} I_{k_x} & \mathbf{0}_{k_x \times k_{z_j}} & -\beta \otimes \mathbf{1}_{1\times J} & -\alpha \otimes \mathbf{1}_{1\times J} \\ \mathbf{0}_{k_{z_j} \times k_x} & I_{k_{z_j}} & -\gamma \otimes \mathbf{1}_{1\times J} & -\rho \otimes \mathbf{1}_{1\times J} \end{pmatrix}, \tag{1.25}$$

$$\Sigma_\delta = E[R(\omega)R(\omega)^{'}] - \delta\delta^{'}, \tag{1.26}$$

$$R(\omega) = \begin{pmatrix} R_1(\omega)' & \dots & R_J(\omega)' \end{pmatrix}^{'}, \; R_j(\omega) = f(\omega)\frac{\partial \sigma_j(\omega)}{\partial \omega_j} - [m_j - \sigma_j(\omega)]\frac{\partial f(\omega)}{\partial \omega_j}, \tag{1.27}$$

$$\delta = (E[\frac{\partial f(\omega)}{\partial \omega_1}m_1]^{'}, \dots, E[\frac{\partial f(\omega)}{\partial \omega_J}m_J]^{'})^{'}, \; \omega_j = \omega \backslash z_{-j} = (x^{'}, z_j^{'}, y_1, \dots, y_J, w_1, \dots, w_J). \tag{1.28}$$

Proof. See the Appendix. □

The efficient GMM estimator $\hat{\theta}^*$ can be obtained by letting $W = \Sigma_g^{-1} = (T\Sigma_\delta T^{'})^{-1}$. Formally, the asymptotic distribution of $\hat{\theta}^*$ is stated in the following proposition.

Proposition 6. *Under the same assumptions as Proposition 5, the efficient GMM estimator, defined by (1.22), has the following asymptotic distribution*

$$\sqrt{n}(\hat{\theta}^* - \theta) \xrightarrow{d} N(0, \Sigma_\theta^*), \tag{1.29}$$

where

$$\Sigma_\theta^* = (A'(T\Sigma_\delta T^{'})^{-1}A)^{-1}. \tag{1.30}$$

A consistent estimator of Σ_θ^* can be obtained by replacing A, T, and Σ_δ in (1.30) with $\hat{A}$, $\hat{T}$, and $\hat{\Sigma}_\delta$, respectively, where $\hat{A}$ is defined as in (1.21), $\hat{T}$ is defined by substituting θ

with its estimator $\hat{\theta}$ in (1.25),

$$\hat{\Sigma}_\delta = \frac{1}{n}\sum_{i=1}^{n}\hat{R}_n(\omega_i)\hat{R}_n(\omega_i)' - \hat{\delta}\hat{\delta}',$$

$$\hat{R}_n(\omega_i) = \Big(\hat{R}_{n1}(\omega_i)' \quad \ldots \quad \hat{R}_{nJ}(\omega_i)'\Big)', \ \hat{R}_{nj}(\omega) = \frac{1}{n-1}\frac{1}{h_n^{k_\omega+1}}\sum_{l\neq i}(m_{ij}-m_{lj})K'(\frac{\omega_i-\omega_j}{h_n}),$$

$$\hat{\delta} = (\hat{E}[\frac{\partial f(\omega)}{\partial \omega_1}m_1]', \ldots, \hat{E}[\frac{\partial f(\omega)}{\partial \omega_J}m_J]')',$$

and $\hat{E}[\frac{\partial f(\omega)}{\partial \omega_j}m_j]$ is defined by (1.19) and (1.20).

1.6 Simulations

In this section I present evidence for the performance of my semiparametric estimator in Section 5. I generate payoff matrices from the random utility model

$$u_{ij} = \beta x_{1i} + \gamma z_{ij} + y_{ij} + \epsilon_{ij}$$
$$v_{ji} = \alpha x_{2i} + w_i + \eta_{ji},$$

where β, γ, and α are scalar parameters, z_{ij}, y_{ij}, x_{2i}, and w_i are generated from the standard normal distribution, respectively, x_{1i} is generated from the uniform distribution on $[0, 1]$. $\epsilon_{ij} = \nu_i + \tilde{\epsilon_{ij}}$ and $\eta_{ij} = \nu_i + \tilde{\eta_{ji}}$, where ν_i is generated from the uniform distribution on $[0, 1]$ and $(\tilde{\epsilon_{ij}}, \tilde{\eta_{ji}})$ is generated from the standard distribution with zero mean and with variance matrix equal to the identity matrix.

In each of 100 simulations, I simulate data for one single market of size $n = 100$, 1000, 2000, 5000, 10000, with $J = 3$ via the student-proposing deferred acceptance algorithm (Gale and Shapley, 1962). When $n = 100$, the college capacities are assumed to be $(15, 20, 13)$. The total number of available seats does not exceed the number of students in the market. The capacities increase proportionally as we increase the sample size n. The estimator is defined as (1.22). The moments I use here are

$$E[m_0 \frac{\partial f(x_1, y_1, \ldots, , y_J)}{\partial x_1}] - \beta \sum_{j=1}^{J} E[m_0 \frac{\partial f(x_1, y_1, \ldots, , y_J)}{\partial y_j}] = 0$$

and for $j = 1, \ldots, J$,

$$E[m_j \frac{\partial f(z_j, y_j)}{\partial z_j}] - \gamma E[m_j \frac{\partial f(z_j, y_j)}{\partial y_j}] = 0$$
$$E[m_j \frac{\partial f(x_2, w)}{\partial x_2}] - \alpha E[m_j \frac{\partial f(x_2, w)}{\partial w}] = 0.$$

Here, to reduce the dimension of covariates in the kernel function to ease the computational burden, for each moment condition, I integrate out "irrelevant" variables in (1.14) and (1.15), and hereby obtain the above moment conditions. I used a Gaussian kernel of order $s = 4$ and bandwidth for each coordinate k

$$h_k = std_k n^{-2/(2k_{cov}+s+2)},$$

where std_k denotes the standard deviation of the sample and k_{cov} denotes the number of coordinates of the density. The bandwidth used here is smaller than the optimal one for estimation of the density to reduce the asymptotic bias of the average derivative estimates (Härdle and Tsybakov, 1993).

Table 1.1 reports the results for $\hat{\beta}$, $\hat{\gamma}$, and $\hat{\alpha}$, the column $\hat{\beta}$ denotes the average over simulations of the value of $\hat{\beta}$, $\bar{\hat{std}}$ denotes the average over simulations of the estimated standard deviations of $\hat{\beta}$, std denotes the empirical standard deviation of $\hat{\beta}$. Analogous results are reported for $\hat{\gamma}$ and $\hat{\alpha}$. It shows that the estimates of the coefficient of the (student, college) specific variable z_{ij}, $\hat{\gamma}$, are quite close to the true value even when $n = 100$. This implies that the limiting model serves as a reasonable approximation of the finite model even with moderate sample size. The estimates for the student-specific attribute x_{2i}, $\hat{\alpha}$, is getting closer to the true value as $n = 2,000$. The estimation of β involves involves $J+1$ covariates,

Table 1.1: Mean and Standard Deviations of Semiparametric Estimator

	$\beta_0 = 0.5$			$\gamma_0 = 0.5$			$\alpha_0 = 0.5$		
n	$\hat{\beta}$	std	$\bar{\hat{std}}$	$\hat{\gamma}$	std	$\bar{\hat{std}}$	$\hat{\alpha}$	std	$\bar{\hat{std}}$
100	-7.47	46.93	15.27	0.51	0.24	0.16	0.03	1.04	0.67
1000	0.49	1.13	0.53	0.50	0.06	0.04	0.36	0.54	0.30
2000	0.53	0.54	0.34	0.50	0.04	0.03	0.49	0.41	0.20
5000	0.51	0.32	0.19	0.50	0.03	0.02	0.50	0.23	0.11
10000	0.49	0.20	0.13	0.50	0.02	0.01	0.51	0.14	0.07

more than the other two parameters γ and α, and the mean of the estimates $\hat{\beta}$ gets close to the true value when $n = 1,000$.

1.7 Conclusion

In this paper, I studied the nonparametric identification and estimation of a many-to-one matching model with non-transferable utility under a price-theoretic framework. The framework proposed in this paper only uses data on matching outcomes, not on specific assignment mechanism or reported preferences. I focused on a single large market in which the number of agents on one side (e.g. students or workers) goes to infinity while the number of agents on the other side (e.g. colleges or firms) remains fixed. My identification results showed that the derivatives of the utility functions and the joint distribution of the two sides' unobserved random terms are both identified. My nonparametric estimator was proposed based on the identification results using Kernel method and the asymptotic properties were established. Based on the identification results, I also proposed a semiparametric estimator that is root-n consistent, asymptotically normal. The simulation results showed that the performance of my semiparametric estimator was good under moderate sample size.

There are several directions for future research. First, in this paper I considered data on one market and assumed that the unobserved heterogeneity is i.i.d. across students. However, if multi-market data is available, the model can be extended by relaxing this assumption and considering college's unobserved characteristics so that students exhibit correlated preference over each college.

Second, the point identification of the utility parameters requires that the excluded regressors have continuous support. Nevertheless, it is quite common that the excluded regressors are discrete. Therefore, it could be important to re-examine the identification results under such alternative assumption on the excluded regressors.

Third, this paper assumed that colleges value each student independently instead of caring about the composition of the incoming class (as a group). While both assumptions can find their applications in the real world, the latter one is more difficult to model. Specifically, extending the model to allow for complementary preference of colleges raises two main challenges. First, in this context, it's natural to consider group stability, a strong solution concept that considers potential deviations by a group of students. Second, there are multiple equilibria even when the number of students is large (Che, Kim, and Kojima, 2019). I am currently working on the solutions to these issues in a separate paper.

Appendix

1.A Student-proposing deferred acceptance algorithm (Gale and Shapley, 1962)

Step 1.

a. Each student applies to her highest ranked college.

(i) Each college with some quota ns_j picks the ns_j highest ranked students among the ones who applied, puts them on the wait list, and rejects the rest.

Step k.

a. Each student rejected at step $(k-1)$ applies to her most preferred college who has not yet rejected her.

(i) The college considers the ns_j highest ranked students among those who applied and on the wait list and puts them on wait list, rejecting the rest.

Stop. The process repeats, and terminates when all students are either matched or have applied to all the colleges they are willing to attend.

1.B Proofs

Before the proof of Lemma 1, I introduce the definition of regular in Azevedo and Leshno (2016), which is the prerequisite for uniqueness. Let $F_{x_i \times z_i \times \epsilon_i \times \eta_i}$ denote the joint distribution of $(x_i, z_i, \epsilon_i, \eta_i)$.

Definition 3. The distribution of student types $F_{x_i \times z_i \times \epsilon_i \times \eta_i}$ is **regular** if the image under $D(\cdot|F_{x_i \times z_i \times \epsilon_i \times \eta_i})$ of the closure of the set

$$\{p \in [0,\, 1]^J :\, D(\cdot|F_{x_i \times z_i \times \epsilon_i \times \eta_i}) \text{ is not continuously differentiable at p}\}$$

has Lebesgue measure 0.

Proof of Lemma 1

Proof. To apply Theorem 1 in Azevedo and Leshno (2016) to prove the lemma, we only need to check that the distribution of student types is regular.

Recall that the limit demand function of college j can be written as

$$\begin{aligned} D_j(p) &= E[m_{ij}(x_i, \epsilon_i, \eta_i, z_{ij}; z_{i,-j}, p)] \\ &= \int \Bigg\{ 1(p_j \leq v^j(z_{ij}, x_i, \eta_{ji})) \\ \times \prod_{j \neq j'} &\Big[1 - 1(u^j(x_i, z_{ij}, \epsilon_{ij}) < u^{j'}(x_i, z_{ij'}, \epsilon_{ij'}))1(p_{j'} < v^{j'}(z_{ij'}, x_i, \eta_{j'i})) \Big] \Bigg\} dF_{X_i \times Y_i} \times F_{\epsilon_i \times \eta_i} \end{aligned}$$

Since $F_{\epsilon_i \times \eta_i}$ is continuously differentiable and $v^j()$ is continuously differentiable in its last argument, the demand function D_j is continuous differentiable in p. □

Proof of Proposition 1

Proof. The proof is similar in spirit to Azevedo and Leshno (2016) and Agarwal and Somaini (2018).

Uniform Convergence of the Demand Function

Firstly, we show that the demand in the finite economy converges uniformly in p to that in the limiting economy. That is, for each $j \in \{1, \ldots, J\}$,

$$\sup_p |D_{n,j}(x^n, z_j^n, \epsilon^n, \eta^n; z_{-j}^n, p) - D_j(p)| \xrightarrow{p} 0. \tag{1.31}$$

It is equivalent to show

$$\sup_p \left|\frac{1}{n}\sum_{i=1}^{n} m_{ij}(x_i, \epsilon_i, \eta_i, z_{ij}; z_{i,-j}, p) - \mathbb{E}[m_{ij}(x_i, \epsilon_i, \eta_i, z_{ij}; z_{i,-j}, p)]\right| \xrightarrow{p} 0,$$

where the expectation is taken with respect to $(x_i, z_i, \epsilon_i, \eta_i)$.

Recall that

$$\begin{aligned} & m_{ij}(x_i, \epsilon_i, \eta_i, z_{ij}; z_{i,-j}, p) \\ =& 1(j \in Ch(x_i, z_{ij}, \eta_i; p) \\ \times & \prod_{j' \neq j} \Big[1 - 1\Big(u^j(x_i, z_{ij}, \epsilon_{ij}) < u^{j'}(x_i, z_{ij'}, \epsilon_{ij'})\Big)1(j' \in Ch(x_i, z_i, \eta_i; p)\Big]. \end{aligned}$$

Since (i) $p \in [0,1]^J$, which is a compact set, (ii) $m_{ij}(x_i, \epsilon_i, \eta_i, z_{ij}; z_{i,-j}, p)$ is continuous at each $p \in [0,1]^J$ with probability one since $v(\cdot)$ is continuous in η_{ji} and η_{ji} has a continuous distribution, (iii) $m_{ij}(x_i, \epsilon_i, \eta_i, z_{ij}; z_{i,-j}, p)$ is dominated by the constant 2, applying the uniform law of large numbers (Newey and McFadden, 1994, Lemma 2.4) proves that (1.31) holds and $D_j(p)$ is continuous in p.

Convergence of the Equilibrium Cutoff

Now that we have the uniform convergence of the demand function, the second part of the proof is to show $p^{(n)*}$ is close enough to p^*.

Let Q_n denote the excess demand in the finite economy

$$Q_n(p; x^n, z^n, \epsilon^n, \eta^n, s) = \|D_n(p; x^n, z^n, \epsilon^n, \eta^n) - s\|,$$

and Q_0 be the excess demand in the limiting economy

$$Q_0(p; s) = \|D(p) - s\|,$$

where D_n and D stacks the demand function of all colleges in the finite economy and limiting economy, respectively. For simplicity of notation, we suppress arguments other than p in the following steps.

Note that the following conditions hold:[12]

(i) $Q_0(p)$ is uniquely minimized by p^* since $Q_0(p^*) = 0$ and p^* is the unique solution by Lemma 1 and $Q_0(p)$ is non-negative;

(ii) Any $p_n \in \mathcal{P}^{(n)}$ minimizes $Q_n(p)$ since $Q_n(p_n) = 0$ and $Q_n(p)$ is non-negative;

(iii) $p \in [0,\ 1]^J$, which is compact;

(iv) $Q_0(p)$ is continuous since $D(p)$ is continuous, which is established in previous steps by uniform law of large numbers;

(v) $\sup_p |Q_n(p) - Q_0(p)| \xrightarrow{p} 0$ by the continuous mapping theorem and (1.31).

Condition (ii) implies that $Q_n(p_n) \leq Q_n(p^*)$. Therefore,

$$\begin{aligned}
& \sup_{p_n \in \mathcal{P}_n} Q_0(p_n) - Q_0(p^*) \\
= & \sup_{p_n \in \mathcal{P}_n} Q_0(p_n) - Q_n(p^*) + Q_n(p^*) - Q_0(p^*) \\
\leq & \sup_{p_n \in \mathcal{P}_n} Q_0(p_n) - Q_n(p_n) + Q_n(p^*) - Q_0(p^*) \\
= & \sup_{p_n \in \mathcal{P}_n} [Q_0(p_n) - Q_n(p_n)] + Q_n(p^*) - Q_0(p^*) \\
\leq & \quad 2 \sup_{p \in [0,1]^J} |Q_0(p_n) - Q_n(p_n)|
\end{aligned}$$

The last inequality is because $\mathcal{P}_n \subset [0, 1]^J$.

[12] The idea of the proof is similar in spirit to the proof of consistency of M-estimators Newey and McFadden (1994).

Thus, for any $\delta > 0$,

$$\mathbb{P}(\sup_{p_n \in \mathcal{P}_n} |Q_0(p_n) - Q_0(p^*)| > \delta) \leq \mathbb{P}(2\sup_{p}|Q_0(p_n) - Q_n(p_n)| > \frac{\delta}{2}) \to 0. \tag{1.32}$$

By conditions (i), (iv) and the fact that $[0,\, 1]^J \cap \{p :\, ||p - p^*|| > \epsilon\}$ is compact, we have

$$\inf_{p\colon ||p-p^*||>\epsilon} Q_o(p) > Q_0(p^*).$$

As a result, for any $\epsilon > 0$, there exists $\delta > 0$ such that

$$\inf_{p\colon ||p-p^*||>\epsilon} Q_o(p) - Q_0(p^*) > \delta,$$

which implies that

$$\sup_{p_n \in \mathcal{P}_n} ||p_n - p^*|| > \epsilon \Rightarrow \inf_{p_n \in \mathcal{P}_n} Q_o(p_n) - Q_0(p^*) > \delta.$$

Hence,

$$\begin{aligned} \mathbb{P}(\sup_{p_n \in \mathcal{P}_n} ||p_n - p^*|| > \epsilon) \qquad & \leq \mathbb{P}(\inf_{p_n \in \mathcal{P}_n} Q_o(p_n) - Q_0(p^*) > \delta) \\ & \leq \mathbb{P}(\sup_{p_n \in \mathcal{P}_n} |Q_o(p_n) - Q_0(p^*)| > \delta) \to 0. \end{aligned}$$

□

Proof of Proposition 2

Proof. For simplicity of notations, assume x and z are scalars. For $k = 1, 2$, let $\omega = (x, z, y^{(k)}, w^{(k)})$. For any $j, j^{'} = 1, \ldots, J$, taking derivaitves of $\sigma_{j^{'}}^{(k)}$ with respect to y_j, w_j, x, and z_j, respectively, one obtains

$$\frac{\partial \sigma_{j'}^{(k)}}{\partial y_j} = -\frac{\partial \Lambda_{j'}(\omega^{(k)})}{\partial \epsilon_j} \tag{1.33}$$

$$\frac{\partial \sigma_{j'}^{(k)}}{\partial w_j} = -\frac{\partial \Lambda_{j'}(\omega^{(k)})}{\partial \eta_j} \tag{1.34}$$

$$\frac{\partial \sigma_{j'}^{(k)}}{\partial x} = -\sum_{j=1}^{J} \frac{\partial \Lambda_{j'}(\omega^{(k)})}{\partial \epsilon_j} u_x^j - \sum_{j=1}^{J} \frac{\partial \Lambda_{j'}(\omega^{(k)})}{\partial \eta_j} v_x^j \tag{1.35}$$

$$\frac{\partial \sigma_{j'}^{(k)}}{\partial z_j} = -\frac{\partial \Lambda_{j'}(\omega^{(k)})}{\partial \epsilon_j} u_{z_j}^j - \frac{\partial \Lambda_{j'}(\omega^{(k)})}{\partial \eta_j} v_{z_j}^j \tag{1.36}$$

Substituting (1.33)-(1.34) into (1.36), one gets

$$\frac{\partial \sigma_{j'}^{(k)}}{\partial x} = \sum_{j=1}^{J} u_x^j \frac{\partial \sigma_{j'}^{(k)}}{\partial y_j} + \sum_{j=1}^{J} v_x^j \frac{\partial \sigma_{j'}^{(k)}}{\partial w_j} \tag{1.37}$$

$$\frac{\partial \sigma_{j'}^{(k)}}{\partial z_j} = u_{z_j}^j \frac{\partial \sigma_{j'}^{(k)}}{\partial y_j} + v_{z_j}^j \frac{\partial \sigma_{j'}^{(k)}}{\partial w_j}. \tag{1.38}$$

Equation (1.37) and (1.38) imply (1.6) in Section 4.2. Denote, as in the statement of the proposition, $\Pi_{x_{(l)}}^j$ to be the matrix formed by replacing the $j-th$ column of matrix Π by the vector $T_{x_{(l)}}$. Let $x_{(l)}$ denote one coordinate of x and $z_{j(l)}$ denote a coordinate of z_j. By the Cramer's rule, for $j = 1, \ldots, J$,

$$\frac{\partial u^j(x, z_j)}{\partial x_{(l)}} = \frac{|\Pi|}{|\Pi_{x_{(l)}}^j|}, \quad \frac{\partial v^j(z_j, x)}{\partial x_{(l)}} = \frac{|\Pi|}{|\Pi_{x_{(l)}}^{j+J}|}.$$

Moreover, for any $j_1, j_2 = 1, \ldots, J$, given that $\frac{\partial \sigma_{j_1}}{\partial y_j}\frac{\partial \sigma_{j_1}}{\partial w_j} - \frac{\partial \sigma_{j_2}}{\partial y_j}\frac{\partial \sigma_{j_2}}{\partial w_j} \neq 0$,

$$\frac{\partial u^j(x, z_j)}{\partial z_j} = \Big(\frac{\partial \sigma_{j_1}}{\partial z_j}\frac{\partial \sigma_{j_2}}{\partial w_j} - \frac{\partial \sigma_{j_2}}{\partial z_j}\frac{\partial \sigma_{j_1}}{\partial w_j}\Big)\Big/\Big(\frac{\partial \sigma_{j_1}}{\partial y_j}\frac{\partial \sigma_{j_2}}{\partial w_j} - \frac{\partial \sigma_{j_2}}{\partial y_j}\frac{\partial \sigma_{j_1}}{\partial w_j}\Big),$$

$$\frac{\partial v^j(z_j, x)}{\partial z_j} = \Big(\frac{\partial \sigma_{j_1}}{\partial z_j}\frac{\partial \sigma_{j_2}}{\partial y_j} - \frac{\partial \sigma_{j_2}}{\partial z_j}\frac{\partial \sigma_{j_1}}{\partial y_j}\Big)\Big/\Big(\frac{\partial \sigma_{j_1}}{\partial y_j}\frac{\partial \sigma_{j_2}}{\partial w_j} - \frac{\partial \sigma_{j_2}}{\partial y_j}\frac{\partial \sigma_{j_1}}{\partial w_j}\Big).$$

$\square$

Proof of Proposition 3

Proof. Wlog, we prove the identification of F when $J = 2$. The argument applies to general cases for any positive integer J. For any j, let $\bar{\epsilon}_j = -u(x, z_j) - y_j$ and $\bar{\eta}_j = p_j - v(z_j, x) - w_j$. Note that $\Lambda_0(\bar{\epsilon}_1, \bar{\epsilon}_2, \bar{\eta}_1, \bar{\eta}_2)$ in (1.7) can be written as a linear combination of joint c.d.f. of subsets of $(\epsilon_1, \epsilon_2, \eta_1, \eta_2)$:

$$
\begin{aligned}
&\Lambda_0(\bar{\epsilon}_1, \bar{\epsilon}_2, \bar{\eta}_1, \bar{\eta}_2)\\
=&\mathbb{P}(\epsilon_1 \leq \bar{\epsilon}_1, \epsilon_2 \leq \bar{\epsilon}_2, \eta_1 \leq \bar{\eta}_1, \eta_2 \leq \bar{\eta}_2) + \mathbb{P}(\epsilon_1 \leq \bar{\epsilon}_1, \epsilon_2 \leq \bar{\epsilon}_2, \eta_1 > \bar{\eta}_1, \eta_2 \leq \bar{\eta}_2)\\
+&\mathbb{P}(\epsilon_1 > \bar{\epsilon}_1, \epsilon_2 \leq \bar{\epsilon}_2, \eta_1 \leq \bar{\eta}_1, \eta_2 \leq \bar{\eta}_2) + \mathbb{P}(\epsilon_1 \leq \bar{\epsilon}_1, \epsilon_2 \leq \bar{\epsilon}_2, \eta_1 \leq \bar{\eta}_1, \eta_2 > \bar{\eta}_2)\\
+&\mathbb{P}(\epsilon_1 \leq \bar{\epsilon}_1, \epsilon_2 > \bar{\epsilon}_2, \eta_1 \leq \bar{\eta}_1, \eta_2 \leq \bar{\eta}_2) + \mathbb{P}(\epsilon_1 \leq \bar{\epsilon}_1, \epsilon_2 \leq \bar{\epsilon}_2, \eta_1 > \bar{\eta}_1, \eta_2 > \bar{\eta}_2)\\
+&\mathbb{P}(\epsilon_1 \leq \bar{\epsilon}_1, \epsilon_2 > \bar{\epsilon}_2, \eta_1 > \bar{\eta}_1, \eta_2 \leq \bar{\eta}_2) + \mathbb{P}(\epsilon_1 > \bar{\epsilon}_1, \epsilon_2 \leq \bar{\epsilon}_2, \eta_1 \leq \bar{\eta}_1, \eta_2 > \bar{\eta}_2)\\
+&\mathbb{P}(\epsilon_1 > \bar{\epsilon}_1, \epsilon_2 > \bar{\epsilon}_2, \eta_1 \leq \bar{\eta}_1, \eta_2 \leq \bar{\eta}_2)\\
=&\mathbb{P}(\epsilon_1 \leq \bar{\epsilon}_1, \epsilon_2 \leq \bar{\epsilon}_2, \eta_1 \leq \bar{\eta}_1, \eta_2 \leq \bar{\eta}_2) + \mathbb{P}(\epsilon_1 \leq \bar{\epsilon}_1, \epsilon_2 \leq \bar{\epsilon}_2, \eta_1 > \bar{\eta}_1)\\
+&\mathbb{P}(\epsilon_1 > \bar{\epsilon}_1, \eta_1 \leq \bar{\eta}_1, \eta_2 \leq \bar{\eta}_2) + \mathbb{P}(\epsilon_2 \leq \bar{\epsilon}_2, \eta_1 \leq \bar{\eta}_1, \eta_2 > \bar{\eta}_2)\\
+&\mathbb{P}(\epsilon_1 \leq \bar{\epsilon}_1, \epsilon_2 > \bar{\epsilon}_2, \eta_2 \leq \bar{\eta}_2)\\
=&F(\bar{\epsilon}_1, \bar{\epsilon}_2, \bar{\eta}_1, \bar{\eta}_2) + F_{\epsilon_1\epsilon_2}(\bar{\epsilon}_1, \bar{\epsilon}_2) - F_{\epsilon_1\epsilon_2\eta_1}(\bar{\epsilon}_1, \bar{\epsilon}_2, \bar{\eta}_1) + F_{\eta_1\eta_2}(\bar{\eta}_1, \bar{\eta}_2) - F_{\epsilon_1\eta_1\eta_2}(\bar{\epsilon}_1, \bar{\eta}_1, \bar{\eta}_2)\\
+&F_{\epsilon_2\eta_1}(\bar{\epsilon}_2, \bar{\eta}_1) - F_{\epsilon_2\eta_1\eta_2}(\bar{\epsilon}_2, \bar{\eta}_1, \bar{\eta}_2) + F_{\epsilon_1\eta_2}(\bar{\epsilon}_1, \bar{\eta}_2) - F_{\epsilon_1\epsilon_2\eta_2}(\bar{\epsilon}_1, \bar{\epsilon}_2, \bar{\eta}_2) \qquad (1.39)
\end{aligned}
$$

It's easy to see that

$$
\begin{aligned}
F_{\epsilon_1\epsilon_2}(\bar{\epsilon}_1, \bar{\epsilon}_2) &= \Lambda_0(\bar{\epsilon}_1, \bar{\epsilon}_2, -\infty, -\infty)\\
F_{\epsilon_1\eta_2}(\bar{\epsilon}_1, \bar{\eta}_2) &= \Lambda_0(\bar{\epsilon}_1, -\infty, -\infty, \bar{\eta}_2)\\
F_{\epsilon_2\eta_1}(\bar{\epsilon}_2, \bar{\eta}_1) &= \Lambda_0(-\infty, \bar{\epsilon}_2, \bar{\eta}_1, -\infty)\\
F_{\eta_1\eta_2}(\bar{\eta}_1, \bar{\eta}_2) &= \Lambda_0(-\infty, -\infty, \bar{\eta}_1, \bar{\eta}_2)
\end{aligned}
$$

Besides,

$$
\begin{aligned}
\Lambda_0(\bar{\epsilon}_1, \bar{\epsilon}_2, \bar{\eta}_1, -\infty) &= \mathbb{P}\Big(\{(\epsilon_1 \le \bar{\epsilon}_1) \cup (\eta_1 \le \bar{\eta}_1)\} \cap (\epsilon_2 \le \bar{\epsilon}_2)\Big) \\
&= \mathbb{P}\Big(\{(\epsilon_1 \le \bar{\epsilon}_1) \cap (\epsilon_2 \le \bar{\epsilon}_2)\} \cap \{(\eta_1 \le \bar{\eta}_1) \cap (\epsilon_2 \le \bar{\epsilon}_2)\}\Big) \\
&= \mathbb{P}\Big((\epsilon_1 \le \bar{\epsilon}_1) \cap (\epsilon_2 \le \bar{\epsilon}_2)\Big) + \mathbb{P}\Big((\eta_1 \le \bar{\eta}_1) \cap (\epsilon_2 \le \bar{\epsilon}_2)\Big) \\
&- \mathbb{P}\Big((\epsilon_1 \le \bar{\epsilon}_1) \cap (\epsilon_2 \le \bar{\epsilon}_2) \cap (\eta_1 \le \bar{\eta}_1)\Big) \\
&= F_{\epsilon_1\epsilon_2}(\bar{\epsilon}_1, \bar{\epsilon}_2) + F_{\epsilon_2\eta_1}(\bar{\epsilon}_2, \bar{\eta}_1) - F_{\epsilon_1\epsilon_2\eta_1}(\bar{\epsilon}_1, \bar{\epsilon}_2, \bar{\eta}_1).
\end{aligned}
$$

Therefore, $F_{\epsilon_1\epsilon_2\eta_1}(\bar{\epsilon}_1, \bar{\epsilon}_2, \bar{\eta}_1) = \Lambda_0(\bar{\epsilon}_1, \bar{\epsilon}_2, -\infty, -\infty) + \Lambda_0(-\infty, \bar{\epsilon}_2, \bar{\eta}_1, -\infty) - \Lambda_0(\bar{\epsilon}_1, \bar{\epsilon}_2, \bar{\eta}_1, -\infty)$ is identified. Similarly, $F_{\epsilon_1\epsilon_2\eta_2}$, $F_{\epsilon_1\eta_1\eta_2}$ and $F_{\epsilon_2\eta_1\eta_2}$ are identified. Then $F(\bar{\epsilon}_1, \bar{\epsilon}_2, \bar{\eta}_1, \bar{\eta}_2)$ is also identified from (1.39). □

Proof of Lemma 2.

Proof. Recall that the conditions (1.12) and (1.13):

$$
\begin{aligned}
\frac{\partial \sigma_j(\omega)}{\partial x} &= \beta \sum_{j'=1}^{J} \frac{\partial \sigma_j(\omega)}{\partial y_{j'}} + \alpha \sum_{j=1}^{J} \frac{\partial \sigma_j(\omega)}{\partial w_{j'}} \\
\frac{\partial \sigma_j(\omega)}{\partial z_j} &= \gamma \frac{\partial \sigma_j(\omega)}{\partial y_j} + \rho \frac{\partial \sigma_j(\omega)}{\partial w_j}.
\end{aligned}
$$

Multiplying the above equations by the density functions $f(\omega)$ and taking expectation on both sides, one get

$$
\begin{aligned}
E[\frac{\partial \sigma_j(\omega)}{\partial x} f(\omega)] &= \beta \sum_{j=1}^{J} E[\frac{\partial \sigma_j(\omega)}{\partial y_j} f(\omega)] + \alpha \sum_{j=1}^{J} E[\frac{\partial \sigma_j(\omega)}{\partial w_j} f(\omega)] \\
E[\frac{\partial \sigma_j(\omega)}{\partial z_j} f(\omega)] &= \gamma E[\frac{\partial \sigma_j(\omega)}{\partial y_j} f(\omega)] + \rho E[\frac{\partial \sigma_j(\omega)}{\partial w_j} f(\omega)].
\end{aligned}
$$

Assumptions 1(ii) and 2(ii) guarantees that σ_j is continuously differentiable in the components of ω for all $\omega \in \bar{\Omega}$, where $\bar{\Omega}$ differs from Ω_0 by a set of measure zero. Also, $E(m_j^2|\omega) = \sigma_j(\omega)$ and therefore is continuous in ω. These conditions, together with Assumption 4, imply that Assumptions 1-3 in Powell et al. (1989) are satisfied. Hence, by Lemma 2.1 in Powell et al. (1989), the above two equations can be written as

$$-2E[\frac{\partial f(\omega)}{\partial x}m_j] = -2\beta\sum_{j=1}^{J}E[\frac{\partial f(\omega)}{\partial y_j}m_j] - 2\alpha\sum_{j=1}^{J}E[\frac{\partial f(\omega)}{\partial w_j}m_j]$$
$$-2E[\frac{\partial f(\omega)}{\partial z_j}m_j] = -2\gamma E[\frac{\partial f(\omega)}{\partial y_j}m_j] - 2\rho E[\frac{\partial f(\omega)}{\partial w_j}m_j],$$

which proves the moment conditions in Lemma 3. $\square$

Before the proof of Proposition 4, we introduce two lemmas. For any $j = 1, \ldots, J$, let $h^j(\omega) = (h_1(\omega), h_2^j(\omega))'$, where $h_1(\omega) = f(\omega)$ is the density of ω and $h_2^j(\omega) = E[m_j|\omega]h_1(\omega)$. For simplicity, denote $\frac{\partial h_1(\omega)}{\partial\omega}$ by $h_{1\omega}$ and denote $\frac{\partial h_2^j(\omega)}{\partial\omega}$ by h_{2w}^j.

Lemma 3. *Let H denote the set of functions h that satisfy Assumption 3(i). Let $\left\|h\right\|_1$ denote the Sobolev norm of order 1. That is, $\|h\|_1 = \max_{l\le 1}\sup_{\omega\in\Omega}\left\|\frac{\partial^l h(x)}{\partial x^l}\right\|$. Define the functional $\varphi_\omega(\cdot)$ on H by*

$$\varphi_\omega(h,\omega) = \frac{\partial(\frac{h_2(\omega)}{h_1(\omega)})}{\partial\omega} = \frac{h_{2\omega}}{h_1} - \frac{h_{1\omega}h_2}{h_1^2}.$$

Assume that for any $j = 1, \ldots, J$, $h^j = (h_1, h_2^j)$ belongs to H and it is such that for $\tau > 0$ and all ω, $h_1(\omega) > \tau$ and $h_2^j(\omega) > \tau$. Then there exists finite $a > 0$ and $\tau_0 > 0$ such that, for all $\tilde{h}^j \in H$ with $\|\tilde{h}^j\|_1 \le \tau_0$ and all ω,

$$\varphi_\omega(h^j + \tilde{h}^j, \omega) - \varphi_\omega(h^j, \omega) = D\varphi_\omega(h^j, \omega; \tilde{h}^j) + R\varphi_\omega(h^j, \omega; \tilde{h}^j),$$

where

$$D\varphi_\omega(h^j, \omega; \tilde{h}^j) = \frac{\tilde{h}_{2\omega}^j h_1 - h_{2\omega}^j\tilde{h}_1 - \tilde{h}_{1\omega}h_2^j - h_{1\omega}\tilde{h}_2^j}{h_1^2} - 2\frac{h_{1\omega}h_2^j\tilde{h}_1}{h_1^3};$$

$$
\begin{aligned}
R\varphi_\omega(h^j,\omega;\tilde{h}^j) = {} & \frac{(h^j_{2\omega}+\tilde{h}^j_{2\omega})(h_1+\tilde{h}_1)-(h_{1\omega}+\tilde{h}_{1\omega})(h^j_2+\tilde{h}^j_2)}{(h_1+\tilde{h}_1)^2} \\
& - \frac{h^j_{2\omega}h_1 - h_{1\omega}h^j_2}{h_1^2} - \frac{\tilde{h}^j_{2\omega}h_1 - h^j_{2\omega}\tilde{h}_1 - \tilde{h}_{1\omega}h^j_2 - h_{1\omega}\tilde{h}^j_2}{h_1^2} \\
& + 2\frac{h_{1\omega}h^j_2\tilde{h}_1}{h_1^3}.
\end{aligned}
$$

Moreover,

$$
\left|D\varphi_\omega(h^j,\omega;\tilde{h}^j)\right| \le a\|\tilde{h}^j\|_1, \ \textit{and} \ \left|R\varphi_\omega(h^j,\omega;\tilde{h}^j)\right| \le a\|\tilde{h}^j\|_1^2.
$$

Proof. Define the functional

$$
\varphi_\omega(h,\omega) = \frac{\partial(\frac{h_2(\omega)}{h_1(\omega)})}{\partial\omega} = \frac{h_{2\omega}}{h_1} - \frac{h_{1\omega}h_2}{h_1^2}.
$$

Then, $\frac{\partial\hat{\sigma}_j(\omega)}{\partial\omega} = \varphi_\omega(\hat{h}^j,\omega)$ and $\frac{\partial\sigma_j(\omega)}{\partial\omega} = \varphi_\omega(h^j,\omega)$.

Let $\tau_0 = \min\{\tau, 1\}$. We first show that for all $\|\tilde{h}^j\|_1 \le \tau_0$,

$$
\varphi_\omega(h^j+\tilde{h}^j,\omega) - \varphi_\omega(h^j,\omega) = D\varphi_\omega(h^j,\omega;\tilde{h}^j) + R\varphi_\omega(h^j,\omega;\tilde{h}^j).
$$

Define

$$
D\varphi_\omega(h^j,\omega;\tilde{h}^j) = \frac{\tilde{h}^j_{2\omega}h_1 - h^j_{2\omega}\tilde{h}_1 - \tilde{h}_{1\omega}h^j_2 - h_{1\omega}\tilde{h}^j_2}{h_1^2} - 2\frac{h_{1\omega}h^j_2\tilde{h}_1}{h_1^3}
$$

and

$$
R\varphi_\omega(h^j,\omega;\tilde{h}^j) = \varphi_\omega(h^j+\tilde{h}^j,\omega) - \varphi_\omega(h^j,\omega) - D\varphi_\omega(h^j,\omega;\tilde{h}^j).
$$

Denote N', N, D', and D by

$$
N' = (h^j_{2\omega}+\tilde{h}^j_{2\omega})(h_1+\tilde{h}_1) - (h_{1\omega}+\tilde{h}_{1\omega})(h^j_2+\tilde{h}^j_2),
$$

$$N = h^j_{2\omega} h_1 - h_{1\omega} h^j_2,$$

$$D^{'} = (h_1 + \tilde{h}_1)^2,$$

$$D = h_1^2.$$

Then

$$\begin{aligned}\varphi_\omega(h^j + \tilde{h}^j, \omega) - \varphi_\omega(h^j, \omega) &= \frac{(h^j_{2\omega} + \tilde{h}^j_{2\omega})(h_1 + \tilde{h}_1) - (h_{1\omega} + \tilde{h}_{1\omega})(h^j_2 + \tilde{h}^j_2)}{(h_1 + \tilde{h}_1)^2} \\ &- \frac{h^j_{2\omega} h_1 - h_{1\omega} h^j_2}{h_1^2} \\ &= \frac{N^{'}}{D^{'}} - \frac{N}{D}.\end{aligned}$$

We employ the following equality used in Matzkin (2015) to derive $R_h \varphi_\omega(h^j, \omega; \tilde{h}^j)$

$$\frac{N^{'}}{D^{'}} - \frac{N}{D} = \frac{N^{'} D - N D^{'}}{D^2} - \frac{(D^{'} - D)(N^{'} D - N D^{'})}{D^{'} D^2}.$$

We also have

$$N^{'} = N + h^j_{2\omega} \tilde{h}_1 + \tilde{h}^j_{2\omega} h_1 - h_{1\omega} \tilde{h}_2 - \tilde{h}_{1\omega} h_2 + R_N,$$

$$D^{'} = D + 2 h_1 \tilde{h}_1 + D_N,$$

where

$$R_N = \tilde{h}^j_{2\omega} \tilde{h}_1 - \tilde{h}_{1\omega} \tilde{h}^j_2 \text{ and } D_N = (\tilde{h}^j_2)^2.$$

Then

$$N' D - N D^{'} = (h^j_{2\omega} \tilde{h}_1 + \tilde{h}^j_{2\omega} h_1 - h_{1\omega} \tilde{h}^j_2 - \tilde{h}_{1\omega} h^j_2) D + R_N D + N \cdot 2 h_1 \tilde{h}_1 + N D_N,$$

$$D^{'} - D = 2 h_1 \tilde{h}^j_2 + D_N,$$

and

$$R\varphi_\omega(h^j,\omega;\tilde{h}^j) = \frac{DR_N - ND_N}{D^2} - \frac{(D^{'} - D)(N'D - ND^{'})}{D^2D^{'}}.$$

It remains to show that for some $a > 0$,

$$\left|D\varphi_\omega(h^j,\omega;\tilde{h}^j)\right| \leq a\|\tilde{h}^j\|_1, \text{ and } \left|R\varphi_\omega(h^j,\omega;\tilde{h}^j)\right| \leq a\|\tilde{h}^j\|_1^2.$$

Let $a = \max\left\{\frac{4\|h^j\|_1}{\tau^2} + \frac{2\|h^j\|_1^2}{\tau^3}, \frac{4\|h^j\|_1^2}{\tau^4} + \frac{(2\|h^j\|_1+1)(8\|h^j\|_1^3+4\|h^j\|_1^2)}{\tau^6}\right\}$

Since $h_1 > \tau$ and $|\tilde{h}_1| \leq \tau$, it follows that $h_1 + \tilde{h}_1 > \tau/2$. It then follows that $D > \tau^2$ and $D^{'} > \tau^2/4$. So

$$\begin{aligned}|D\varphi_\omega(h^j,\omega;\tilde{h}^j)| &\leq (\frac{4\|h^j\|_1}{\tau^2} + \frac{2\|h^j\|_1^2}{\tau^3})\|\tilde{h}^j\|_1 \\ &\leq a\|\tilde{h}^j\|_1.\end{aligned}$$

Also, for all $\tilde{h}^j$ with $\|h^j\|_1 < \tau$,

$$\left|D\right| \leq \|h^j\|_1^2,\ |R_N| \leq 2\|\tilde{h}^j\|_1^2,\ |N| \leq 2\|h^j\|_1^2,\ |D_N| \leq \|\tilde{h}^j\|_1^2,$$

$$\left|D^{'} - D\right| \leq (2\|h^j\|_1+1)\|\tilde{h}^j\|_1,$$

$$|N'D - ND^{'}| \leq (8\|h^j\|_1^3+4\|h^j\|_1^2)\|\tilde{h}^j\|_1^2.$$

Therefore,

$$\begin{aligned}|R\varphi_\omega(h^j,\omega;\tilde{h}^j)| &\leq \Big(\frac{4\|h^j\|_1^2}{\tau^4} + \frac{(2\|h^j\|_1+1)(8\|h^j\|_1^3+4\|h^j\|_1^2)}{\tau^6}\Big)\|\tilde{h}^j\|_1^2 \\ &\leq a\|\tilde{h}^j\|_1^2.\end{aligned}$$

$\square$

Proof of Proposition 4

Proof. Let $(y,w)=(y_1,\dots,y_J,w_1,\dots,w_J)$. Let $h=(h_1,\dots,h_J)$. Define the functionals

$$\Psi_\Pi(h,\omega^{(1)},\omega^{(2)})=\begin{bmatrix}\varphi_{(y,w)}(h^1,\omega^{(1)})\\ \vdots\\ \varphi_{(y,w)}(h^J,\omega^{(1)})\\ \varphi_{(y,w)}(h^1,\omega^{(2)})\\ \vdots\\ \varphi_{(y,w)}(h^J,\omega^{(2)})\end{bmatrix},$$

$$\Psi_{x_{(l)}}(h,\omega^{(1)},\omega^{(2)})=\begin{bmatrix}\varphi_{x_{(l)}}(h^1,\omega^{(1)}) & \cdots & \varphi_{x_{(l)}}(h^J,\omega^{(1)}) & \varphi_{x_{(l)}}(h^1,\omega^{(2)}) & \cdots & \varphi_{x_{(l)}}(h^J,\omega^{(2)})\end{bmatrix},$$

and

$$\Xi(h,\omega^{(1)},\omega^{(2)})=[\Psi_\Pi(h,\omega^{(1)},\omega^{(2)})]^{-1}\Psi_{x_{(l)}}(h,\omega^{(1)},\omega^{(2)}).$$

For simplicity, we suppress $(\omega^{(1)},\omega^{(2)})$ in the arguments in the following steps. By Lemma 3, it can shown that there exists finite $a_1>0$ and $\tau_1>0$, a linear funcitonal $D\Xi$, and a functional $R\Xi$ such that, when $\|h\|_1<\tau_1$,

$$\Xi(h+\tilde h)-\Xi(h)=D\Xi(h;\tilde h)+R\Xi(h;\tilde h),$$

where

$$\begin{aligned}D\Xi(h;\tilde h)=&[\Psi_\Pi(h)]^{-1}D\Psi_\Pi(h;\tilde h)[\Psi_\Pi(h)]^{-1}\Psi_{x_{(l)}}(h)\\ &+[\Psi_\Pi(h)]^{-1}D\Psi_{x_{(l)}}(h;\tilde h),\end{aligned}$$

and

$$\|D\Xi(h;\tilde h)\|_1\le a_1\|\tilde h\|_1,\ \text{and}\ \|R\Xi(h;\tilde h)\|_1\le a_1\|\tilde h\|_1^2.$$

When $\hat{h} = h + \tilde{h}$, Assumptions 3(i)(iii)(iv) together with Lemma B.3 in Newey (1994) imply that $\|\hat{h}-h\|_1 = O_p(\sqrt{\ln(n)/(nh_n^{k_\omega+2})} + h_n^s)$. By Assumption 3(iv), $\sqrt{nh_n^{k_\omega+2}}(\sqrt{\ln(n)/(nh_n^{k_\omega+2})} + h_n^s)^2 \to 0$. Therefore, since $|R\varphi_\omega(h,\omega;\hat{h}-h)| \leq a\|\hat{h}-h\|_1^2$, it follows that $\sqrt{nh_n^{k_\omega+2}}R\varphi_\omega(h,\omega;\hat{h}-h) = o_p(1)$. Therefore, $\sqrt{nh_n^{k_\omega+2}}R\Xi(h;\tilde{h}) = o_p(1)$ and

$$\begin{aligned}
&\sqrt{nh_n^{k_\omega+2}}\Big(\Xi(\hat{h}) - \Xi(h)\Big) \\
&= \sqrt{nh_n^{k_\omega+2}}D\Xi(h;\tilde{h}) + \sqrt{nh_n^{k_\omega+2}}R\Xi(h;\tilde{h}) \\
&= \sqrt{nh_n^{k_\omega+2}}D\Xi(h;\tilde{h}) + o_p(1).
\end{aligned}$$

For any $j, j' = 1, \ldots, J$, define the following $2J \times 2J$ matrices: r_1, r_{1,y_j}, r_{1,w_j}, $r_2^{j'}$, $r_{2,y_j}^{j'}$, and $r_{2,w_j}^{j'}$. Let r_1 denote the matrix whose $j-th$ row is equal to $-\frac{\frac{\partial h_2^j(\omega^{(1)})}{\partial(y,w)}h_1(\omega^{(1)})+2\frac{\partial h_1(\omega^{(1)})}{\partial(y,w)}h_2^j(\omega^{(1)})}{[h_1(\omega^{(1)})]^3}$ and $(j+J)-th$ row is equal to $-\frac{\frac{\partial h_2^j(\omega^{(2)})}{\partial(y,w)}h_1(\omega^{(2)})+2\frac{\partial h_1(\omega^{(2)})}{\partial(y,w)}h_2^j(\omega^{(2)})}{[h_1(\omega^{(2)})]^3}$. Let r_{1,y_j} denote the matrix whose $j-th$ column is equal to $-\begin{bmatrix}\frac{h_2^1(\omega^{(1)})}{[h_1(\omega^{(1)})]^2} & \cdots & \frac{h_2^J(\omega^{(1)})}{[h_1(\omega^{(1)})]^2} & \frac{h_2^1(\omega^{(2)})}{[h_1(\omega^{(2)})]^2} & \cdots & \frac{h_2^J(\omega^{(2)})}{[h_1(\omega^{(2)})]^2}\end{bmatrix}'$ and the other entries are equal to zeros. Let r_{1,w_j} denote the matrix whose $(j+J)-th$ column is equal to $-\begin{bmatrix}\frac{h_2^1(\omega^{(1)})}{[h_1(\omega^{(1)})]^2} & \cdots & \frac{h_2^J(\omega^{(1)})}{[h_1(\omega^{(1)})]^2} & \frac{h_2^1(\omega^{(2)})}{[h_1(\omega^{(2)})]^2} & \cdots & \frac{h_2^J(\omega^{(2)})}{[h_1(\omega^{(2)})]^2}\end{bmatrix}'$ and the other entries are equal to zeros. Let $r_2^{j'}$ denote the matrix whose $j'-th$ row is equal to $-\frac{\frac{\partial h_1(\omega^{(1)})}{\partial(y,w)}}{[h_1(\omega^{(1)})]^2}$, $(j'+J)-th$ row is equal to $-\frac{\frac{\partial h_1(\omega^{(2)})}{\partial(y,w)}}{[h_1(\omega^{(2)})]^2}$, and the other entries are equal to zeros. Let $r_{2,y_j}^{j'}$ denote the matrix whose (j',j) element is equal to $\frac{1}{h_1(\omega^{(1)})}$, $(j'+J,j)$ element is equal to $\frac{1}{h_1(\omega^{(2)})}$, and the other entries are equal to zeros. Let $r_{2,w_j}^{j'}$ denote the matrix whose $(j',j+J)$ element is equal to $\frac{1}{h_1(\omega^{(1)})}$, $(j'+J,j+J)$ element is equal to $\frac{1}{h_1(\omega^{(2)})}$, and the other entries are equal to zeros. We also define the following $2J \times 1$ vectors: $\tilde{r}_1$, $\tilde{r}_{1,x_{(l)}}$, $\tilde{r}_2^{j'}$, and $\tilde{r}_{2,x_{(l)}}^{j'}$. Let $\tilde{r}_1$ be the vector whose $j-th$ element is $-\frac{\frac{\partial h_2^j(\omega^{(1)})}{\partial x_{(l)}}h_1(\omega^{(1)})+2\frac{\partial h_1(\omega^{(1)})}{\partial x_{(l)}}h_2^j(\omega^{(1)})}{[h_1(\omega^{(1)})]^3}$ and $(j+J)-th$ element is $-\frac{\frac{\partial h_2^j(\omega^{(2)})}{\partial x_{(l)}}h_1(\omega^{(2)})+2\frac{\partial h_1(\omega^{(2)})}{\partial x_{(l)}}h_2^j(\omega^{(2)})}{[h_1(\omega^{(2)})]^3}$. Let $\tilde{r}_{1,x_{(l)}}$ be the vector whose $j-th$ element is equal to $-\frac{h_2^j(\omega^{(1)})}{[h_1(\omega^{(1)})]^2}$ and $(j+J)-th$ element is equal to $-\frac{h_2^j(\omega^{(2)})}{[h_1(\omega^{(2)})]^2}$. Let $\tilde{r}_2^{j'}$ denote the vector whose $j'-th$ and $(j'+J)-th$ elements are equal to $-\frac{\frac{\partial h_1(\omega^{(1)})}{\partial x_{(l)}}}{[h_1(\omega^{(1)})]^2}$ and $-\frac{\frac{\partial h_1(\omega^{(2)})}{\partial x_{(l)}}}{[h_1(\omega^{(2)})]^2}$, respectively, and the other entries are equal to zeros. Let $\tilde{r}_{2,x_{(l)}}^{j'}$ be the vector whose $j'-th$ and $(j'+J)-th$

elements are equal to $\frac{1}{h_1(\omega^{(1)})}$ and $\frac{1}{h_1(\omega^{(2)})}$, respectively, and the other entries are equal to zeros.

For any $j, j^{'} = 1, \ldots, J$, define the following $2J \times 1$ vectors

$$\Gamma_1 = \Pi^{-1} r_1 \Pi^{-1} T_{x(l)} + \Pi^{-1} \tilde{r}_1, \quad \Gamma_{1,y_j} = \Pi^{-1} r^{j^{'}}_{1,y_j} \Pi^{-1} T_{x(l)}, \tag{1.40}$$

$$\Gamma_{1,w_j} = \Pi^{-1} r^{j^{'}}_{1,w_j} \Pi^{-1} T_{x(l)}, \quad , \Gamma_{1,x_{(l)}} = \Pi^{-1} \tilde{r}_{1,x_{(l)}}, \tag{1.41}$$

$$\Gamma^{j^{'}}_{2} = \Pi^{-1} r^{j^{'}}_{2} \Pi^{-1} T_{x(l)} + \Pi^{-1} \tilde{r}^{j^{'}}_{2}, \quad , \Gamma^{j^{'}}_{2,y_j} = \Pi^{-1} r^{j^{'}}_{2,y_j} \Pi^{-1} T_{x(l)}, \tag{1.42}$$

$$\Gamma^{j^{'}}_{2,w_j} = \Pi^{-1} r^{j^{'}}_{2,w_j} \Pi^{-1} T_{x(l)}, \quad , \Gamma^{j^{'}}_{2,x_{(l)}} = \Pi^{-1} \tilde{r}^{j^{'}}_{2,x_{(l)}}. \tag{1.43}$$

Further define

$$\bar{\Gamma}_1 = \begin{bmatrix} \Gamma_1 & \Gamma_{1,y_1} & \cdots & \Gamma_{1,y_J} & \Gamma_{1,w_1} & \cdots & \Gamma_{1,w_J} & \Gamma_{1,x_{(l)}} \end{bmatrix}, \text{ and} \tag{1.44}$$

$$\bar{\Gamma}^j_2 = \begin{bmatrix} \Gamma^j_2 & \Gamma^j_{2,y_1} & \cdots & \Gamma^j_{2,y_J} & \Gamma^j_{2,w_1} & \cdots & \Gamma^j_{2,w_J} & \Gamma^j_{2,x_{(l)}} \end{bmatrix}. \tag{1.45}$$

Define the $2J \times (2J+2)(J+1)$ matrix

$$\Gamma = \begin{bmatrix} \bar{\Gamma}_1 & \bar{\Gamma}^1_2 & \cdots & \bar{\Gamma}^J_2 \end{bmatrix}. \tag{1.46}$$

Define $\tilde{K}(\omega) = (K(\omega), \frac{\partial K(\omega)}{\partial y_1}, \ldots, \frac{\partial K(\omega)}{\partial y_J}, \frac{\partial K(\omega)}{\partial w_1}, \ldots, \frac{\partial K(\omega)}{\partial w_J}, \frac{\partial K(\omega)}{\partial x_{(l)}})'$.

Let $\Sigma(\omega) = \begin{bmatrix} 1 & \sigma_1(\omega) & \cdots & \cdots & \sigma_J(\omega) \\ \sigma_1(\omega) & \sigma_1(\omega) & 0 & \cdots & 0 \\ \vdots & 0 & \sigma_2(\omega) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_J(\omega) & 0 & 0 & \cdots & \sigma_J(\omega) \end{bmatrix}$.

Let

$$V = \Gamma[\Sigma(\omega) \otimes \int \tilde{K}(\tilde{\omega}) \tilde{K}(\tilde{\omega})^{'} d\tilde{\omega}]\Gamma^{'}. \tag{1.47}$$

and let

$$\hat{V} = \hat{\pi}(\omega)'[\hat{\Sigma}(\omega) \otimes \int \tilde{K}(\tilde{\omega})\tilde{K}(\tilde{\omega})' d\tilde{\omega}]\hat{\pi}(\omega)$$

be an estimator for V, obtained by substituting h_1 and h_2^j by $\hat{h}_1$ and $\hat{h^j}_2$ respectively in the definitions of r_1, r_{1,y_j}, r_{1,w_j}, $r_2^{j'}$, $r_{2,y_j}^{j'}$, $r_{2,w_j}^{j'}$, $\tilde{r}_1$, $\tilde{r}_{1,x_{(l)}}$, $\tilde{r}_2^{j'}$, and $\tilde{r}_{2,x_{(l)}}^{j'}$, respectively, and substituting σ_j by $\hat{\sigma}_j$, defined in (1.9), in the definition of Σ.

Assumptions 3(i)(ii) imply that Γ is bounded and continuous almost everywhere and zero outside a compact set Ω. Assumption 2(ii) guarantees that σ_j is continuous a.e.. so $\Sigma(\omega)$ is continuous a.e. In addition, for $\epsilon > 0$, $\sup_{\|\eta\|<\epsilon} (1 + \sigma_j(\omega + \eta))f(\omega + \eta) < \infty$ by Assumption 3(ii). Hence, Assumption 5.1 in Newey (1994) is satisfied. Assumptions 3(i)(iii) imply that Assumptions K, H, and Y in Newey (1994) are also satisfied. Therefore, it follows by Assumption 3(iv) and Lemma 3 in Newey (1994), by letting in that lemma, $k_1 = k_\omega$, $k_2 = 0$, and $l = 1$, that

$$\sqrt{nh_n^{k_\omega+2}} D\Xi(h; \tilde{h}) \xrightarrow{d} N(0, V),$$

where V is defined in (1.47). Therefore,

$$\sqrt{nh_n^{k_\omega+2}} \Big(\Xi(\hat{h}) - \Xi(h)\Big) \xrightarrow{d} N(0, V). \tag{1.48}$$

To show $\hat{V}$ is a consistent estimator for V, define the functional Υ corresponding to Γ. Following similar arguments as those on $\varphi_\omega(h^j, \omega)$ in Lemma 3, it can be shown that there exists a constant c and functionals $D\Upsilon$ and $R\Upsilon$ such that

$$\begin{aligned} |\Upsilon(\hat{h}, \omega) - \Upsilon(h, \omega)| &= |D\Upsilon(h; \hat{h} - h) + R\Upsilon(h; \hat{h} - h)| \\ &\leq |D\Upsilon(h; \hat{h} - h)| + |R\Upsilon(h; \hat{h} - h)| \\ &\leq c\|\hat{h} - h\|_1 + c\|\hat{h} - h\|_1^2. \end{aligned}$$

Therefore, it follows from $\|\hat{h} - h\|_1 = o_p(1)$ that $|\Upsilon(\hat{h}, \omega) - \Upsilon(h, \omega)| = o_p(1)$. This together with the fact that $|\hat{\sigma}_j(\omega) - \sigma_j(\omega)| = o_p(1)$ implies that $\hat{V} \xrightarrow{p} V$. □

Proof of Proposition 5

Proof. We first of all develop asymptotic normality for $\sqrt{n}\hat{g}_n(\theta_0)$. For this purpose, we first derive the asymptotic distribution for all the average derivative estimators in $\hat{g}_n$, denoted by $\delta = (\delta_1^{'}, \dots, \delta_J^{'})^{'}$. Here $\delta_j = E[\frac{\partial f(\omega)}{\partial \omega_j} m_j]$ is the vector of expectation terms involved in the moment condition $g_j(\theta)$, where $\omega_j = \omega \backslash z_{-j} = (x^{'}, z_j^{'}, y_1, \dots, y_J, w_1, \dots, w_J)$.

By Assumptions 2(ii), 4, and 5, applying Theorem 3.3 in Powell et al. (1989), we have

$$\sqrt{n}(\hat{\delta} - \delta) \xrightarrow{d} N(0, \Sigma_\delta),$$

where $\Sigma_\delta = E[R(\omega)R(\omega)^{'}] - \delta\delta^{'}$ and $R(\omega) = \begin{pmatrix} R_1(\omega)^{'} & \dots & R_J(\omega)^{'} \end{pmatrix}^{'}$ with $R_j(\omega) = f(\omega)\frac{\partial \sigma_j(\omega)}{\partial \omega_j} - [m_j - \sigma_j(\omega)]\frac{\partial f(\omega)}{\partial \omega_j}$.

Let $T = \begin{pmatrix} T_1 & 0 & \cdots & 0 \\ 0 & T_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & T_J \end{pmatrix}$ with $T_j(\theta) = \begin{pmatrix} I_{k_x} & \mathbf{0}_{k_x \times k_{z_j}} & -\beta \otimes \mathbf{1}_{1\times J} & -\alpha \otimes \mathbf{1}_{1\times J} \\ \mathbf{0}_{k_{z_j} \times k_x} & I_{k_{z_j}} & -\gamma \otimes \mathbf{1}_{1\times J} & -\rho \otimes \mathbf{1}_{1\times J} \end{pmatrix}$.
Then $g(\theta) = T(\theta)\delta$. By Delta Method,

$$\sqrt{n}(\hat{g}_n(\theta) - g(\theta)) \xrightarrow{d} N(0, \Sigma_g),$$

where $\Sigma_g = T\Sigma_\delta T^{'}$.

Next we derive asymptotic distribution for my GMM estimator $\hat{\theta}$.

Obviously, $g_j(\omega, \theta) == \begin{pmatrix} m_j \frac{\partial f(\omega)}{\partial x} - \beta \sum_{j=1}^{J} m_j \frac{\partial f(\omega)}{\partial y_j} - \alpha \sum_{j=1}^{J} m_j \frac{\partial f(\omega)}{\partial w_j} \\ m_j \frac{\partial f(\omega)}{\partial z_j} - \gamma m_j \frac{\partial f(\omega)}{\partial y_j} - \rho m_j \frac{\partial f(\omega)}{\partial w_j} \end{pmatrix}$ is continuously differentiable in each $\theta \in \Theta$. Assumption 6 (iii) guarantees that $E[\left\| g(\omega, \theta) \right\|^2] < \infty$ and $E[\sup_{\theta \in \Theta} \left\| \nabla_\theta g(\omega, \theta) \right\|] < \infty$. In addition, given Assumption 6(i)(ii), all regularity conditions for the asymptotic normality of GMM estimator are satisfied. Hence, we have

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \Sigma_\theta),$$

where Σ_θ is given by (1.24). $\square$

Chapter 2

Estimating Preferences in the U.S. College Admissions under a Matching Framework

2.1 Introduction

In this paper, I estimate the preferences of the students and colleges in the U.S. college admissions and conduct counterfactual analysis, based on the theoretical results in Chapter 1. The U.S. college admissions market is a decentralized market in the sense that students directly apply to and receive offers from colleges without coordination of any centralized clearinghouse. The solution I propose this paper is to recover the preferences only based on matching outcomes.

To implement the empirical analysis, I apply the proposed semiparametric estimation strategy in Chapter 1 to the data set on U.S. college admissions, primarily constructed from High School Longitudinal Study of 2009 (HSLS:09) and the Integrated Postsecondary Education Data System (IPEDS). My estimation results indicate that net price (i.e., tuition minus financial aid/scholarship) of attending college has a negative impact on student's preferences over colleges while student's SES (e.g., family income and parental education) has a positive effect on a student's inclination to attend colleges. Furthermore, a student's ability as measured by test scores and a math ability index leads to a higher evaluation by colleges. In the empirical estimation, the majority of effects are significant at the one percent level and large in magnitude.

My first counterfactual experiment evaluates the welfare gains and losses in the decentralized market relative to the centralized market under deferred acceptance (DA) algorithm, a setting that does not involve strategic behaviors. In a decentralized market, students directly approach multiple colleges, and after receiving offers, decide which one to accept. In a centralized market, students express preferences in the rank order lists, colleges submit admission criteria, which determine students' priority, after which an algorithm automatically matches students and colleges.[1] The U.S. is one of the few countries that has adopted

[1] Theoretically, if the students are rational, these two markets should generate the same results when the number of students is large due to the uniqueness of equilibrium shown by Azevedo and Leshno (2016). However, this is not the case in reality since students may have different degrees of "sophistication". For example, Hoxby and Avery (2013) discussed a puzzling phenomenon that a large number of low-income high achievers do not apply to any selective college even when the cost of attending selective college is lower

a decentralized college admissions process while many countries use centralized mechanisms to match students with colleges. I find that overall, average welfare is similar in the two markets. However, compared with the centralized market, average welfare for high-SES and high-ability students is greater in the decentralized market by \$5,289 and \$11,767, respectively, while average welfare for first-generation college students and low-ability students is lower in the decentralized market by \$1,333 and \$1,578, respectively. These patterns are consistent with the previous findings that high-SES students benefit from the uncoordinated market relative to a coordinated one (Hemphill et al., 2009; Abdulkadiroğlu, Agarwal, and Pathak, 2017).

The second counterfactual experiment accesses the welfare change of eliminating the SAT score from the evaluation process. My finding suggests that average welfare loss of eliminating SAT is small; however, students with highly-educated parents and high SAT scores suffer non-negligible welfare losses.

Contributions to the Literature

This paper relates to studies on the college market in an equilibrium framework. Epple, Romano, and Sieg (2006) and Epple et al. (2017) develop and estimate general equilibrium models that characterize equilibrium policies. In Epple et al. (2017), by choosing admission and financial aid policies, private college maximizes college quality, which depends on the average student ability and educational resources provided to students, while public college maximizes the aggregate achievement of in-state students. Fu (2014) develops and estimates a structural subgame perfect Nash equilibrium (SPNE) model, where students face uncertainty and bear application costs when making application decisions. My model considers matching between students and colleges based on their preferences, taking tuition and financial aid policies as given. It simplifies the application decisions of students and focuses on the college's admission decision and students' enrollment decisions. An implicit assumption

than that of non-selective colleges. In the counterfactual analysis, I quantify the differences between the two markets caused by student's different degrees of sophistication and ability to make decisions.

is that the application cost is zero. This assumption is in line with Epple, Romano, and Sieg (2006) and Epple et al. (2017). My model adds to the literature by allowing for unobserved heterogeneity in college preference and by not specifying the joint distribution of colleges' unobserved error terms, while Epple, Romano, and Sieg (2006), Fu (2014), and Epple et al. (2017) assume away the unobserved college preference. My model, for example, covers the scenario in which colleges value the quality of students' essays or interviews, which are not observed by the econometrician. Also, the preference over students' essays may be correlated across colleges.

This paper also contributes to the literature that studies the welfare effects of various allocation mechanisms (Abdulkadiroğlu and Sönmez, 2003; Che and Koh, 2016; Abdulkadiroğlu, Agarwal, and Pathak, 2017; Liu, Wan, and Yang, 2018). It provides empirical evidence on whether a decentralized system will cause welfare gains or losses, and how gains or losses are distributed across participants. The literature on such evidence is remarkably small, partly because it is challenging to develop a framework of decentralized market where the allocation process is hard to observe. Abdulkadiroğlu, Agarwal, and Pathak (2017) and Liu, Wan, and Yang (2018) also quantify the welfare implications of decentralized markets. Abdulkadiroğlu, Agarwal, and Pathak (2017) use data on the reform of the New York City public high school assignment system and find that the transition from an uncoordinated market to a coordinated market significantly improved average student welfare. Liu, Wan, and Yang (2018) use data from a ride-sharing platform in China and find that centralized algorithms can improve welfare by increasing both the match quality and the number of matches. This paper adds to this literature by investigating the welfare consequences of the U.S. college admissions system.

2.2 The Specification

To implement the empirical analysis, I consider the following specification

$$u_{ij} = \beta x_{1i} + \gamma z_{ij} - d_{ij} + \epsilon_{ij}$$

$$v_{ji} = \alpha x_{2i} + SAT_i + \eta_{ji}.$$

Here, u_{ij} is the utility of student i over college j, and v_{ji} is the utility of college j over student i. x_{i1} includes student i's family income and parental education. z_{ij} includes net tuition paid by student i to college j as a cost measure and college j's SAT 25th percentile score as a quality measure.[2] x_{2i} includes high school GPA and a math ability index. Specifically, the math ability index is a score constructed by the High School Longitudinal Study of 2009 (HSLS:09) and represents the probability that a student would pass a given proficiency level.[3] I use this variable to capture student's logic thinking ability and to serve as a proxy for ability to make decisions. The variables d_{ij} and SAT_i act as exclusive regressors; d_{ij} is the distance between student i's home (as proxied by her high school) and college j. Distance is assumed to influence the demand of a student on colleges without affecting their choice sets. SAT_i leads to variations in the choice set of students without changing student preferences.[4] The idea is similar to exploiting exclusion restrictions in simultaneous equation model and isolating variation in each side of the market.[5] The coefficient of d_{ij} is normalized to -1 so that β and γ are measured in terms of willingness to travel. The normalization of the coefficient of SAT_i to 1 implies that α measures the importance of the other ability

[2] The most common measure of quality in the literature is the median SAT score (Dale and Krueger, 2002; Brewer, Eide, and Ehrenberg, 1999; Terry Long, 2004). IPEDS provides the 25th and 75th percentile SAT score instead of the median SAT score. My results are robust to using the 75th percentile SAT as the quality measure.

[3] There are seven levels of proficiency and the levels are hierarchical in the sense that mastery of a higher level typically implies proficiency at the lower levels. I use the highest level in the analysis. The results are robust when using other levels.

[4] Here the exclusive regressor SAT_i does not have to vary across colleges because all the observables in college's preferences are student-specific.

[5] Similarly, in his analysis of the medical match, Agarwal (2015) used the birth and medical school location, median MCAT score, and log NIH funding of graduating medical school as exclusive regressors.

determinants relative to SAT score. The coefficients of d_{ij} and SAT_i are scale normalized because in general, the units of utility are not identified. ϵ_{ij} refers to the unobserved taste shock in student i's preference over college j. The error term η_{ji} refers to the unobserved taste shock in college j's preference over student i such as essays, recommendations, and extracurriculars. Here the joint distribution of $(\epsilon_{i1}, \ldots, \epsilon_{iJ}, \eta_{1i,} \ldots, \eta_{Ji})$ is still assumed to be fully nonparametric.

2.3 Descriptive Statistics

The main data set is constructed using the High School Longitudinal Study of 2009 (HSLS:09) and the Integrated Postsecondary Education Data System (IPEDS), both of which are collected by the National Center for Education Statistics (NCES). HSLS:09 contains detailed information for students who first enrolled in college in 2013 on applications, admissions, financial aid, and high school transcript. IPEDS provides detailed information of the postsecondary institutions over years and I use the characteristics of the colleges in 2012, when students would have been applying/choosing a college. I also use Common Core of Data (CCD) and Private School Universe Survey (PSS) to obtain zip codes of high schools. Table 2.1 and 2.2 shows the summary statistics of the colleges and students, respectively.[6]

I define net tuition, quality, and distance for each potential (student, college) pair in the following ways. First, the student is assigned the in-state tuition if she is a resident of the same state as the college; otherwise, she is assigned out-of-state tuition. To construct net price, I subtract the reported or estimated amount of the scholarship or grant that the student receives from the tuition.[7] Second, I use 25th percentile SAT math score and 25th percentile SAT reading score of the college's student body to proxy the college's quality.

[6] Since the tables involve restricted-use data, as requested by NCES, in all the tables, all sample size numbers are rounded to the nearest ten.

[7] The amount of scholarships and grants was subtracted if it was reported; otherwise, if the student was offered Pell grant, the average amount of Pell grant ($3,579) in 2013 was subtracted. In addition, if the student was offered work-study funding, the average amount ($1,669) in 2013 was subtracted. As we do not observe whether the reported amount of scholarships and grants was offered by federal financial aid or the enrolled college, I assume that the reported amount does not vary with college for each student.

Table 2.1: Summary Statistics of College Characteristics

	Mean	Sd	Median	N
Public (0/1)	0.4	0.49	0	870
Elite (0/1)	0.09	0.29	0	870
In-state tuition	20,685	13,254	22,266	870
Out-of-state tuition	25,027	9,837	24,468	870
SAT 25th percentile score				
Critical reading	495.51	63.51	480	760
Math	507.2	67.42	490	760
Region				
South (0/1)	0.33	0.47	0	870
West (0/1)	0.12	0.33	0	870
Northeast (0/1)	0.28	0.45	0	870
Midwest (0/1)	0.27	0.45	0	870

The definition of elite colleges follows Fu (2014). Based on U.S. News and World Report from 2012, the top 30 private universities, top 20 liberal arts colleges, and the top 30 public universities are considered as elite colleges. As requested by NCES, in all the tables, all sample size numbers are rounded to the nearest ten.

To reflect the assortative matching between student's interest of study and college quality, instead of using the sum of the two 25th percentile scores, I weight each 25th percentile score by student's relative advantage of the subject. Specifically, the weight on college SAT math score is defined as the ratio of the student's high school math GPA percentile rank over the sum of the student's high school math GPA percentile rank and English GPA percentile rank, and similar for the weight on college SAT reading score. Third, distance is constructed using the zip codes (centroid) of student's high school and the college. HSLS:09 also provides which college the student is enrolled as of November 1, 2013. I use this information to generate the matching outcome of the market.

Because the first step of my semiparametric method is kernel estimation of the density function, it suffers from the curse of dimensionality as a nonparametric estimation approach. In the equilibrium model, the matching outcome of any college depends on all the colleges' characteristics. Therefore, the dimension of covariates is a multiple of the number of colleges, leading to severe curse of dimensionality when there are many colleges. I reduce the dimensionality of the covariates in the kernel function in three ways. First, I treat non-selective colleges as the outside option and only consider selective colleges as alternatives. Second,

Table 2.2: Summary Statistics of Student Characteristics

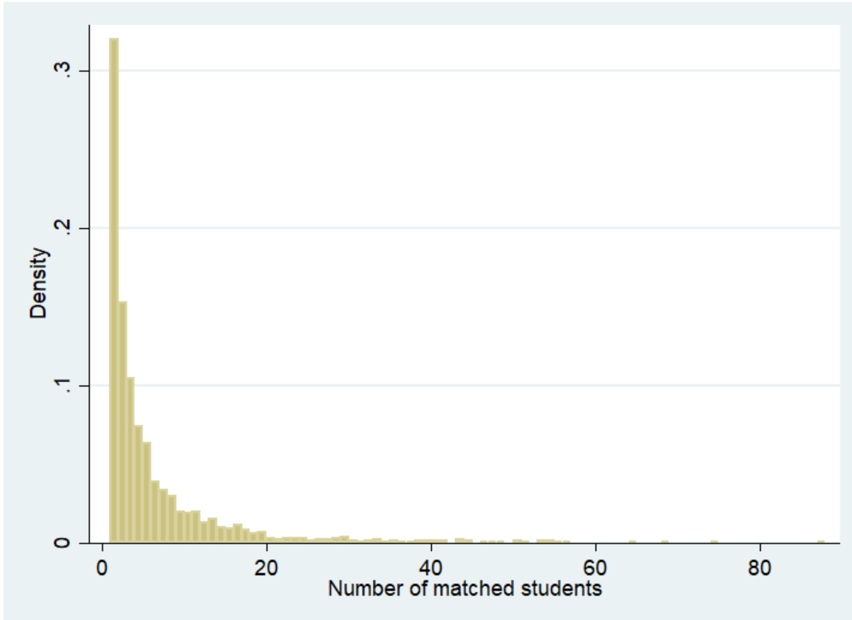
	Mean	Sd	N
Female (0/1)	0.53	0.5	13,940
Family income	89,947	71,700	13,390
Parent's years of schooling	15.99	3.53	13,780
Composite SAT	1,030.46	200.83	7,240
High school GPA	2.99	0.69	13,290
Math ability index	0.04	0.09	13,280
Number of postsecondary institutions applied	3.11	2.82	13,940
Enrolled (0/1)	0.87	0.34	13,940
Enrolled in selective college (0/1)	0.48	0.5	13,940
Scholarship	11,546	18,656	6,010
Work study (0/1)	0.24	0.42	10,070
Pell grant (0/1)	0.5	0.5	10,330

Sample includes the students who applied to at least one college. Composite SAT is defined as the sum of actual math score and critical reading score where available; otherwise was predicted using the student's high school math GPA, English GPA, gender, race, and ethnicity. Math ability index is defined as the math proficiency probability score, which represents the probability that a student would pass a given proficiency level. Selectivity of college is defined according to Carnegie Classification of Institutions of Higher Education. As requested by NCES, in all the tables, all sample size numbers are rounded to the nearest ten.

in the moment conditions, I integrate out "irrelevant" variables in the same way as in the simulations. Third, I aggregate the colleges into groups and treat these groups as alternatives. The aggregation is necessary also because a large fraction of colleges in the data are matched to very few students (the distribution of the number of matched students across colleges is shown in Figure 2.1). I aggregate colleges by public/private, elite/nonelite, and region.[8] I obtain 16 groups when the region is divided by south/west/northeast/midwest. I also consider an alternative finer division of region: if the total enrollment in a state is over 200 in the data, I treat this state as a regional group; I then aggregate the remaining states according to south/west/northeast/midwest as four other regional groups. In this way, I obtain 59 groups.

[8]The definition of elite colleges follows Fu (2014). Based on U.S. News and World Report from 2012, the top 30 private universities, top 20 liberal arts colleges, and the top 30 public universities are considered as elite colleges.

Figure 2.1: The distribution of college's number of matched students



2.4 Estimation Results

Table 2.3 shows the estimation results for each aggregation scheme. Under the aggregation scheme with 16 college bins, all the variables except for college SAT 25th percentile score are significant at the one percent level. Recall that the coefficients of a student's preferences are measured in terms of willingness to travel. In magnitude, an increase of $\$1,000$ in net price will lead to a decrease in a student's utility that is equivalent to attending university 90 miles further away from home – about 43% of the mean distance traveled by students in the data.[9] An increase of $\$1,000$ in family income will lead to an increase in a student's utility that is equivalent to attending university 67 miles closer to home – about 32% of the mean distance traveled by students in the data. One additional year of parental schooling increases student's utility by an amount that is equivalent to attending college 1,320 miles closer to home, six times the mean of distance traveled by students in the data. As for the college's utility, the coefficients are measured relative to SAT score. A one-point increase in GPA (i.e., 25% increase relative to the full score) is equivalent to an increase of SAT score by 295

[9] As a reference, distance to actual assignments has mean 207 miles, median 85 miles, and standard deviation 385 miles in the data.

Table 2.3: Preference Parameters

	16 colleges		59 colleges	
	Coefficients	Std. Error	Coefficients	Std. Error
Student's utility				
Net price	-0.090***	0.006	-0.044***	0.003
College 25th percentile SAT score	-0.079	0.462	0.634*	0.362
Family Income	0.067***	0.002	–	–
Parent's years of schooling	1320***	57	–	–
College's utility				
High school GPA	295***	37	60**	23
Math ability index	5040***	326	3510***	322
N	11,680		11,680	

Significance levels indicated by $*p < .10$; $**p < .05$; $***p < .01$. Since the tables involve restricted-use data, as requested by NCES, in all the tables, all sample size numbers are rounded to the nearest ten.

points (i.e., 18% increase relative to the full score). An increase in the probability of passing a math proficiency test by 10% is equivalent to an increase of SAT score by 504 points (i.e., 32% increase relative to the full score). Using the alternative way of aggregating the colleges into 59 groups, I obtain estimates that show similar patterns, although the magnitude of the coefficients is smaller than that in the first aggregation scheme. It is worth noting that in the second aggregation scheme, college quality now has a significantly positive impact on student's utility. A 100-point increase of college SAT 25th percentile score leads to an increase in student's utility that is equivalent to attending college 63 miles closer to home – about 30% of the mean distance traveled by the students. Estimates for student-specific attributes in the second scheme are not calculated because they involve estimating kernel functions with 60 covariates, an exercise which is not computationally feasible given the sample size.

2.5 Welfare Analysis

To compute welfare measures and conduct counterfactual analysis, I first estimate the distribution of the unobserved terms. Although I have proven the nonparametric identification of the joint distribution of the unobserved terms in previous section, to facilitate estimation, here I make some parametric assumptions. Assume $\epsilon_{ij} = \sigma_\epsilon \tilde{\epsilon}_{ij}$, where $\tilde{\epsilon}_{ij}$ follows i.i.d. extreme type I distribution, and $\eta_{ji} = \lambda_i + \tilde{\eta}_{ji}$, where λ_i follows the normal distribution with zero mean and variance σ_λ^2, and $\tilde{\eta}_{ji}$ follows the standard normal distribution.[10] I further assume that $\tilde{\epsilon}_{ij}$, λ_i, and $\tilde{\eta}_{ji}$ are independent. I estimate $(\sigma_\epsilon, \sigma_\lambda)$ by minimizing the distance between the observed probabilities and the predicted probabilities of being matched with each college, treating the cutoffs as auxiliary parameters.

2.5.1 Decentralized vs. Centralized Markets

Because individual student utility is not directly observed, I compute the expectation of the utility for each student given by

$$E[u_{im(i)}|x_{1i}, x_{2i}, z_{ij}, d_{ij}, SAT_i].$$

For the counterfactual centralized market, the expectation is estimated based on the predicted matching outcomes under the Deferred Acceptance (DA) algorithm using simulated ϵ_i and η_i.[11] Average student welfare is defined as the average of the expected utilities that students derive from their assignment (indirect utilities). The DA algorithm requires that both students and colleges submit their rank order list or priority index as inputs. The algorithm will obtain a stable matching after several iterations and will assign to each student at most one college. The centralized market under DA algorithm provides a benchmark where students do not have to engage in strategic behavior. Therefore, the ability to make decisions, as proxied by math ability index, would not play a role in the centralized market.

For the decentralized market, the expectation is calculated based on the predicted matching outcomes given estimated cutoffs using simulated ϵ_i and η_i. Table 2.4 displays the differ-

[10]In most of the literature on logit discrete choice, σ_ϵ is normalized to 1 as a scale normalization. In the model, since I already assume the coefficient of distance to be -1, I do not make any normalization assumption on σ_ϵ.

[11]See appendix for a description of the DA algorithm.

ence in student welfare between the decentralized market and the centralized market for all students and for subgroups by gender, race, ethnicity, SES, and ability. The first and second columns report the difference in average student welfare measured in miles and in dollars, respectively. A positive value means higher welfare in the decentralized market than in the centralized one. The third column shows the total welfare change for the entire cohort (1.66 million students) who took SAT in 2012. Overall, average student welfare is similar between the decentralized market and the centralized market. Welfare change differs by race and ethnicity. Relative to the centralized market, Asians gain by 523 miles in the decentralized market while Blacks and Hispanics lose by 174 miles and 98 miles; welfare change is much smaller for whites than for non-whites.[12] These differences by race or ethnicity may be due to SES, culture, and other social norms. Regarding SES, students whose parents both have bachelor's degrees and who themselves have a high SAT score considerably gain (by 476 miles and 1,059 miles) while students with no parent ever going to college and with low SAT score lose (by 120 miles and 142 miles) in the decentralized market relative to the centralized market. These patterns are consistent with the previous findings that high-SES students benefit from the uncoordinated market relative to a coordinated one (Hemphill et al., 2009; Abdulkadiroğlu, Agarwal, and Pathak, 2017).

In Table 6, I turn to a multivariate analysis to address the importance of demographics, SES, and high school assistance in college preparation. Column (1) in Table 6 includes basic demographics such as gender, race, and ethnicity. The results are consistent with the descriptive statistics in Table 2.4. In column (2), I add indicators for whether mother or father has a bachelor's degree and a measure of family income. All the SES variables have significantly positive impacts on student's welfare in the decentralized market, relative to the centralized DA counterfactual. This result might indicate that parents who attended college are more familiar with the application process and can help their children make more sophisticated decisions or improve their application relative to those who did not. Alternatively,

[12] The actual distance traveled in the data has mean 207 miles, median 85 miles, and standard deviation 385 miles.

Table 2.4: Difference in Welfare: Decentralized - Centralized (DA)

	Change in average welfare		Change of the entire cohort
Unit:	(miles)	(dollars)	(billion dollars)
All	49	544	0.90
Female	-19	-211	-0.19
Male	126	1,400	1.07
Asian	523	5,811	1.25
Black	-174	-1,933	-0.45
Hispanic	-98	-1,089	-0.25
White	11	122	0.15
Low income	-78	-867	-0.40
High income	175	1,944	2.29
Parents with bachelor's degree			
Both	476	5,289	1.84
Only mother	-21	-233	-0.15
Only father	48	533	0.29
Neither	-120	-1,333	-1.11
High ability (SAT>1200)	1,059	11,767	6.25
Low ability (SAT<=1200)	-142	-1,578	-1.78

Low income equals to one if the student's family was at/above or below 185% of the 2012 poverty threshold, as set forth by the U.S. Census Bureau; and zero, otherwise. Both family income and household size are considered when calculating whether a family is at/above or below 185% of the poverty threshold. High income equals to one if the student's family income in 2012 was above the median in the data; and zero otherwise.

high-SES students have access to better high schools and more resources to facilitate their college application process than low-SES students. Furthermore, when the SES variables are included, the difference in welfare change between Black (Hispanic) and White students decline by 43% (82%) in absolute value, implying that SES variables explain a considerate fraction of the race and ethnicity effects. In column (3), I add two variables measuring the assistance provided by the high school with regard to college preparation — whether the high school's counseling staffs spent over 50% of working hours in assisting students with college readiness, selection, and applications, and the fraction of 11th and 12th grade students in the high school who were helped with selecting colleges. Although I cannot directly observe whether a particular student is offered counseling or information, the higher the percentage at the high school who are offered counseling service, the more likely it is that any particular student would have received help directly or benefited indirectly from these services through spillovers from their peers. I find that only the assistance with college selection is significantly correlated with the welfare difference. In column (4), by adding interaction between parental education and high school assistance, I find that mother's education and high school assistance are complementary to each other. This finding implies that college-educated parents could help their child make better use of high school assistance.

2.5.2 Eliminating SAT

An on-going debate in higher education has been about how much weight SAT and ACT scores should be given in the admission process. A growing number of colleges, including some highly selective colleges such as the university of Chicago, no longer mandate SAT or ACT score submissions in college admission process.[13] While proposals for dropping these standardized test stands typically on reducing costs for low-income families, its potential consequences to other groups of students merits examination as well.

I compare average student welfare by simulations under the DA mechanisms, with and without SAT score in the admission process. Table 2.6 displays welfare change induced by eliminating SAT score from the admission process. The first and second columns in Table 2.6 report the change in average student welfare measured in miles and dollars. The third column shows total welfare change of the cohort (1.66 million students) who took SAT in 2012. A negative value implies a welfare loss by eliminating SAT. Overall, average welfare loss is small. But the change differs by group. High-ability and high-SES students suffer a welfare loss when SAT is not required in the admissions process and total welfare loss for these advantaged groups amounts to 0.5 to 2 billion dollars. On the one hand, my finding supports the argument of the policy's advocates that colleges would attract and enroll more underrepresented minority and low-income students by pursuing an optional SAT admissions

[13] A recent news can be found at http://www.chicagotribune.com/news/local/breaking/ct-university-chicago-sat-act-20180614-story.html.

Table 2.5: Welfare Change by Demographics, SES, and Information

Dep: $W^{de} - W^{DA}$	(1)	(2)	(3)	(4)
Female	-141.1***	-123.9***	-127.5***	-128.6***
	(16.9)	(16.5)	(18.1)	(18.1)
Asian	507.3***	470.2***	493.0***	494.4***
	(25.2)	(24.7)	(27.0)	(26.9)
Black	-206.5***	-118.5***	-129.5***	-127.1***
	(24.5)	(24.1)	(27.7)	(27.7)
Hisp	-119.6***	-21.8	-24.9	-28.3
	(24.7)	(24.3)	(26.9)	(26.9)
Other	-101.2	-37.8	-36.7	-40.1
	(86.4)	(84.2)	(93.6)	(93.5)
Mother w/ bachelor's degree		151.8***	155.2***	10.5
		(18.8)	(20.7)	(42.2)
Father w/ bachelor's degree		233.7***	243.4***	150.6***
		(20.0)	(22.1)	(44.4)
Family income		0.001***	0.001***	0.001***
		(0.0001)	(0.0001)	(0.0001)
HS counseling on college			24.1	-26.1
			(32.0)	(48.9)
HS college selection help			1.623***	0.067
			(0.313)	(0.426)
HS counseling on college * mother w/ bachelor's degree				62.1
				(69.5)
HS college selection help * mother w/ bachelor's degree				2.540***
				(0.684)
HS counseling on college * father w/ bachelor's degree				27.9
				(69.8)
HS college selection help * father w/ bachelor's degree				1.596**
				(0.705)
Constant	103.6***	-180.3***	-250.8***	-162.6***
	(13.9)	(18.4)	(24.6)	(29.1)
Observations	11,610	11,610	9,980	9,980
R-squared	0.053	0.102	0.108	0.111

Significance levels indicated by $*p < .10$; $**p < .05$; $***p < .01$. Since the tables involve restricted-use data, as requested by NCES, in all the tables, all sample size numbers are rounded to the nearest ten. HS counseling on college preparation equals to one if counseling staffs in the high school spent over 50% of working time on college readiness/selection/apply; and zero, otherwise. HS help with college selection is defined by the percent of 11th/12th grade students in the high school who were helped with selecting colleges.

Table 2.6: Change in Welfare by Eliminating SAT

	Change in average welfare		Change of the entire cohort
Unit:	(miles)	(dollars)	(billion dollars)
All	-33	-367	-0.61
Female	-16	-178	-0.16
Male	-53	-589	-0.45
Asian	-205	-2,278	-0.49
Black	169	1,878	0.44
Hispanic	112	1,244	0.29
White	-48	-533	-0.66
Low income	79	878	0.41
High income	-75	-833	-0.98
Parents with bachelor's degree			
Both	-255	-2,833	-0.99
Only mother	-45	-500	-0.32
Only father	-67	-744	-0.41
Neither	80	889	0.74
High ability (SAT>1200)	-625	-6,944	-2.0
Low ability (SAT<=1200)	59	656	0.90

Low income equals to one if the student's family was at/above or below 185% of the 2012 poverty threshold, as set forth by the U.S. Census Bureau; and zero, otherwise. Both family income and household size are considered when calculating whether a family is at/above or below 185% of the poverty threshold. High income equals to one if the student's family income in 2012 was above the median in the data; and zero otherwise.

policy. On the other hand, my finding shows a more comprehensive picture by evaluating the cost on students from high-SES family. Therefore, the impact of the policy is inconclusive, pointing to a direction for future research.

2.6 Conclusion

In this paper, I analyzed the U.S. college admissions and estimated the utility parameters of students and colleges under a many-to-one matching framework. By doing so, I uncovered the role of student's family background, college costs, and college quality in student's enrollment decision. In addition, I estimated how student's SAT score and high school gpa determine college's admission decision. Based on the estimation results, I implemented and quantified the welfare effects of two counterfactual experiments: i) experiment where a decentralized market is switched to a centralized market; ii) experiment where the SAT score is eliminated from the admission process. Furthermore, this framework can be used to evaluate the effects of various policies such as expanding college capacity and increasing financial aid.

To further improve the model, I plan to work on the following three aspects. First, it is necessary to provide evidence for the validity of the excluded regressors used in the paper (i.e., distance between the student and the college and SAT score).

Second, I plan to investigate different specifications including interactions and discrete random variables to consider important factors such as race and public/private college.

Third, at the current stage I only compared the student welfare between the decentralized market and the centralized market through the channel of student's degree of sophistication. However, there are still some important channels that need to be explored.

Chapter 3

Three Generations of Changing Gender Patterns of Schooling in China

3.1 Introduction

China has distinguished itself over the past several decades by its extremely rapid economic growth and the accompanying sharp rise in literacy and in living standards. However, it is also notable for the strong son preference that has led to one of the most imbalanced sex ratios in the world. The most recent data puts the ratio at 1.15 boys to girls at birth and 1.17 boys to girls from 0-14 years old (The World Factbook, 2018). Son preference has been a pronounced phenomenon for generations and is evident in several other countries as well. (See Hesketh and Xing, 2006 for a discussion detailing sex ratios in Asia.)

Son preference has been attributed to numerous underlying cultural norms Das Gupta et al. (2003) but is likely also due, at least in part, to economic factors. Parents who may anticipate a need for old age support may prefer sons who not only have had a greater earning potential than daughters, but who, given cultural norms, are more likely to live near parents than are daughters who traditionally have lived nearer their husband's family.

The strong son preference in China is manifested prior to birth in selective abortion and other means of skewing the sex ratio at birth, and subsequently through higher mortality rates for newborn daughters relative to sons. It can also manifest itself as children grow, with fewer resources devoted to daughters than to sons. Studies have shown worse health outcomes for girls over boys (Pande, 2003; Li, Zhu, and Feldman, 2004; Mishra, Roy, and Retherford, 2004), less financial support (Lei et al., 2017), and less investment in schooling (Kingdon, 2002; Wang, 2005; Song, Appleton, and Knight, 2006).

In addition to differences by gender, we focus our attention on differences in schooling investments between rural and urban areas. Not only are economic conditions different across regions, implying differing returns to education, but we might also expect the strength of son preference and other attitudes towards sons and daughters to vary differentially between the two areas as well.

In this paper we use data from the China Health and Retirement Longitudinal Study (CHARLS) to analyze schooling attainment for three generations of Chinese men and women, giving us an unusually long view of changes in investments. While CHARLS is a panel study of individuals ages 45 or older and their spouses, the survey also collects information on the parents and children of these respondents. These generation data provide us the opportunity to examine the time path of schooling investments for individual families and to control for differences *within* families. We can look at intergenerational correlations in schooling attainment for a full three generational dynasty.

In our work we look not just at time trends in completed schooling but attempt to identify the importance of two key policies central to Chinese economic growth: compulsory schooling laws and the enactment of China's one child policy. The effects of these policies are identified through the variation in timing for the roll-outs and the degree of enforcement across geographic regions.

We find large differences in schooling by both region (urban or rural) and gender, but these differences have decreased substantially over time–particularly those differences between the schooling attainment of males and females in urban areas. The gains in schooling for females are large enough that, among the youngest in our sample, women have more years of schooling, on average, than do young men. With respect to the public policy levers, we find that compulsory schooling laws contributed significantly to the rise in schooling, and did so to a greater extent in rural areas, and particularly for females in these areas. Furthermore, the one child policy has had a significant effect on schooling for both boys and girls, likely as parents increase the investment in children rather than attempt to divide that investment across siblings. The lower total cost associated with raising one child rather than several, also allows greater resources to be directed towards education. However, contrary to expectations, we do not find differing effects of the one child policy for educational outcomes for males and females.

In the following section we provide some background on son preference and on schooling attainment. In section three we introduce our data set along with some descriptive statistics. Section four provides our formal regression analysis and a final section offers our conclusions.

3.2 Background

3.2.1 Overview

China's economic growth has indeed been rapid, averaging close to 10 percent per year from 1989 to the present. Going back even further, the peak annual growth rate hit 19.3 percent in 1970 (numbers based on public data from the World Bank). This modernizing and growing economy has meant more resources to finance education and more job opportunities for educated individuals, particularly for women. A more educated populace, in turn, itself contributes to economic growth. Another artifact of modernization, improved health and the increasing life expectancy, accompanying economic growth provides for longer working life and a longer length of time over which to reap returns to an investment in education. And finally, modernization can bring with it changes in attitudes or preferences, particularly the tradition of son preference, leading potentially to smaller biases in the distribution of educational investments.

In addition to the effect of rising economic fortunes in China and increased employment opportunities for women on schooling investments, educational attainment in China was also likely affected by public policy. Compulsory schooling laws introduced by the government beginning in 1986 mandated attendance through grade nine. These mandates would be expected to have had a larger effect on schooling for girls than boys, since girls started at lower levels of education. In providing a larger positive impact on female educational levels, these laws ought to have reduced the difference in the schooling attainment of boys and girls. Similarly, one might expect the effects of compulsory schooling to be larger in rural areas where schooling levels were initially low.

Finally, the one-child policy may also have led not just to greater schooling investment overall, but to a decline in the gap between the education of boys and girls. There are several reasons to posit such an effect. First, as highlighted by Almond, Li, and Meng (2014), with a limit on the number of children a family can have, parents with the strongest son preferences may choose selective abortion or other means to ensure the birth of a son. Thus, those families who do have a daughter would be less averse to investing in a girl and thus more likely to provide her with greater levels of schooling. Second, with a single child, parents have an incentive to invest heavily in that child regardless of gender, and even if the "returns" to investing in a daughter are smaller than those of investing in a son, the investment in schooling will likely yield positive returns. And finally, with fewer children to support, parents can invest more in the schooling of all children. But daughters, starting at a lower average level of schooling, have more potential for gain once financial constraints are eased.

3.2.2 Schooling in China

In 1986 China instituted compulsory schooling requirements, mandating nine years of schooling for all children. Despite the laws, enrollment beyond primary grades remained far from universal for some time. The China Education Research Network provides statistics indicating that in 1990, approximately 75 percent of those finishing primary school "graduated" to middle school.[1] While enrollments increased rapidly in the coming years, the children in our sample were born before 1991 and thus attended primary schools well-before middle school attendance was ubiquitous.

Unsurprisingly, schooling in rural areas has been far below that in urban areas (Connelly and Zheng, 2003) and several factors likely come into play in explaining that difference. In rural areas the opportunity cost of sending a child to school is likely larger because the child could provide labor for the family. Conversely, the returns to schooling are lower in rural areas so the cost of foregoing a middle school education is lower. The "cost" of getting to school in terms of distance traveled / difficulty is also larger in rural areas, reducing attendance (Li and Liu, 2014). And the burden of paying associated school fees and the cost of books may be felt more strongly in rural areas where incomes are lower.[2]

[1] `http://www.edu.cn/gai_kuang_495/20100121/t20100121_441886.shtml`.

[2] Reforms enacted in 2001 through 2006 puts limits on the amount of tuition and fees that could be

3.3 Data

3.3.1 Description of CHARLS

Our data come from the Chinese Health and Retirement Longitudinal Study or CHARLS. CHARLS is a longitudinal survey that is nationally representative of the non-institutionalized Chinese population 45 years old or older and their spouses. The first wave of the survey was fielded in 2011 with follow-up waves in 2013 and 2015 and continuing on a biennial basis. CHARLS is part of a set of "sister surveys" established across a large number of countries, with the respondent populations in all cases, focused on those approaching retirement. These related studies include the Health and Retirement Study in the United States (HRS), the Japanese Survey of Aging and Retirement (JSTAR), the Korean Longitudinal Study of Aging (KLOSA), the English Longitudinal Study of Ageing (ELSA), and the Survey of Health, Aging, and Retirement in Europe (SHARE) among others.[3] The surveys interview both respondents and their spouses obtaining information on income, wealth, health and family relationships.

The initial interview round of CHARLS, undertaken in 2001, surveyed 17,587 individuals in 10,257 households. There are 8,436 males and 9,151 females in this initial waves, with the larger number of women due to differential mortality at older ages.[4] We draw our data primarily from the second wave of the survey in which a total of 18,605 respondents were interviewed.

Our study is focused on trends in schooling and the role of public policies instituted in the later part of the last century in affecting that change. The key policies we analyze are the one child policy and compulsory schooling laws, both of which were implemented

charged and allowed for financial assistance. See Chyi and Zhou (2014) for a discussion of the impact of these reforms.

[3]Other related surveys are the Irish Longitudinal Study on Ageing (TILDA), the Mexican Health and Aging Study (MHAS), the Longitudinal Aging Study in India (LASI), and the Study on Global Ageing and Adult Health (SAGE).

[4]Despite the skewed sex ratio at birth currently, the sex ratio at birth in 1962 was estimated to be 107 boys for every 100 girls–approximately what is expected naturally. (World Bank, 2018)

approximately 30 to 40 years ago. To this end, we construct our analytic sample to focus primarily on the children of our respondents. With parents averaging 60 years old when the survey began, their children were born in precisely the time period most affected by the changes. In fact, the average birth year of the children in our sample is 1977, just prior to the establishment of the one child policy. In addition to these children, we also include the respondents themselves as well as the parents of the respondents (the grandparents of those on whom we focus). While many of these grandparents have already died, the survey collected schooling information and birth year for deceased grandparents so we can thus include them in our analysis of schooling attainments over time.

The use of these data, for three generations of the same family, is a key way in which our study differs from others examining schooling in China. Not only do we have schooling for three generations in the same family, but at the child level, we have schooling for (typically) all children in the family.[5] We can thus look at differences in schooling attainment holding constant family fixed effects. In doing so when focusing on the child generation, we are implicitly examining differences between brothers and sisters within the same family. We can thus assess the importance of gender, holding constant factors such as familial resources or attitudes towards schooling–measures which may be correlated with educational attainment and with the gender composition of the family-leading to biased results.

Both the parents and grandparents in our sample (i.e. the respondents and spouses in CHARLS and their own parents) are well beyond the age at which individuals are likely to be enrolled in school, so the initial reports of educational attainment are an accurate measure of completed schooling. However, at the start of the survey, a substantial fraction of children of the respondents are young enough that they are still accumulating years of schooling. We therefore impose an age cutoff for children, and choose age 22. Furthermore, to ensure that we have as large a sample as possible of those in this generation who are 22 years old or older, we use observations on schooling as measured in the second wave of the

[5] We limit our sample to children 22 years old or older so in some families we will omit younger siblings.

survey–the most recent data available. In doing so, we are providing these children with the potential for two additional years to age and to finish their education relative to schooling as reported in wave 1. Our age cutoff of 22 is thus based on age in wave 2, corresponding to the year 2013. These "children" thus were born in 1991 or earlier. Because our top education category is "some college or more" even those children who are still attending college at age 22 will be denoted as having achieved some college, so their "final" level of schooling in our classification scheme will not change in subsequent waves. It is unlikely that many of those 22 years old or older who have not previously attended college will return to school later in the survey. (Nor will those age 22 or older return to complete primary school, middle school, or high school.) With these restrictions we are left with a sample of 9,751 families with 27,306 children. Excluding families with zero household weight leaves 26,798 children in 9,558 families.

3.3.2 Descriptive Characteristics of Sample

Descriptive statistics for our sample are presented in table 3.1. Here we use one observation per family and use values of time-varying variables as measured in 2013. Because of the large differences between urban and rural areas in important measures such as schooling, income, and the enforcement of the one-child policy, we also report the means separately by the urban/rural status of the respondent household. The two rightmost columns of the table report these means. For this table wherein we measure variables on the household level, we define rural or urban based on the location of the primary respondent at the time of the interview. We note that given China's rapid industrialization, this location may be different from the location of children or grandparents. It may also differ from the location of the respondent at the time he or she was born or the time the children were born. Later in the paper, when analyzing schooling on an individual level, we use the hukou of the individual's birth for each generation.

Seventy-two percent of our total sample is married; the mean birth year for the male

Table 3.1: Weighted Summary Statistics of Family Characteristics

	All *N= 9,558**		Urban *N=3,680*		Rural *N=5,600*	
	Mean	SE	Mean	SE	Mean	SE
Married couple	0.72	(0.005)	0.72	(0.007)	0.73	(0.006)
Birth year of father	1952	(0.109)	1952	(0.179)	1952	(0.138)
Birth year of mother	1953	(0.109)	1953	(0.173)	1953	(0.142)
Number of Children	2.94	(0.016)	2.62	(0.025)	3.23	(0.021)
Only child	0.16	(0.004)	0.24	(0.007)	0.08	(0.004)
Number of Sons	1.53	(0.011)	1.33	(0.017)	1.71	(0.015)
Number of Daughters	1.41	(0.012)	1.29	(0.020)	1.52	(0.017)
Birth year of oldest child	1976	(0.108)	1976	(0.174)	1975	(0.141)
Birth year of youngest child	1981	(0.080)	1981	(0.131)	1981	(0.101)
Family income	35,694	(1011)	49,576	(1210)	24,411	(1467)
Years of Schooling						
Grandfather (father's side)	2.02	(0.040)	2.65	(0.074)	1.47	(0.042)
Grandmother (father's side)	0.765	(0.026)	1.17	(0.052)	0.39	(0.022)
Grandfather (mother's side)	1.93	(0.038)	2.49	(0.071)	1.45	(0.040)
Grandmother (mother's side)	0.66	(0.024)	1.03	(0.050)	0.34	(0.020)
Father	6.80	(0.053)	8.09	(0.089)	5.70	(0.062)
Mother	4.11	(0.052)	5.82	(0.091)	2.68	(0.053)
Sons (within family mean)	9.80	(0.044)	11.23	(0.070)	8.61	(0.052)
Daughters (within family mean)	9.00	(0.054)	10.93	(0.079)	7.35	(0.064)

*Numbers differ across columns and variables due to missing values on some measures. Rural status is missing for 278 households.

respondents in the family (the fathers in our generational approach) is 1952, and for the females (mothers) it is 1953. The mothers in our sample were thus, on average, 26 years old in 1979 when the one child policy was established. Unsurprisingly given the sampling frame, the birth years of respondents are identical across urban and rural regions.

The average number of children for the families in our sample is 2.94, a surprise to many given the recent attention to China's one child policy. Consistent with the well-documented unequal gender ratios, there are more sons than daughters–1.53 to 1.41. Families with only one child are surprisingly rare, comprising just 16 percent of families in our sample. (This fraction rises to 18 percent when including families with children under 22, demonstrating the expected rise over time in the prevalence of one child families.)

It is when looking at these variables—variables pertaining to family size and composition—we see large differences across rural and urban areas, providing our first indication that son preference and investments in children might vary across regions. The average family size in urban areas is 2.62 children compared to 3.23 in rural areas, consistent with the greater costs of raising children in urban areas and the negative correlation typically found between income and family size across countries. Similarly, only children are far more common in urban areas, 0.24 versus 0.08, where not only are children more costly, but the one child policy was more strictly enforced.

When looking at the gender of children, the son-bias appears to be far larger in rural areas than urban areas. In urban areas there are 1.33 sons and 1.29 daughters, on average, for a sex ratio of males to females of 1.031—actually a bit below what is considered normal—compared to 1.71 sons and 1.52 daughters in rural areas with an implied sex ratio of 1.125 in rural areas.

We also see the expected differences in income across regions with household income in urban areas being approximately twice that of rural areas.[6]

There has undoubtedly been a sharp rise in educational levels in China. To examine the changes in schooling over time we stacked the data for all three generations of family members to provide a person level (rather than family level) data set. Each respondent-couple contributes two observations for themselves, four observations for their parents / parents-in-law, and one observation for each child. An unmarried respondent contributes an observation for herself, one for each of her parents, and one for each of her children. While the respondent-based sample is population representative of individuals of the targeted cohort, we note that this expanded person level sample is not population representative for the older

[6]While nearly all our data come from the publicly available data on the CHARLS website, our income measure is that developed by the HRS Harmonization project and available from `https://g2aging.org/`.

and younger cohorts. Nonetheless, we believe these data provide important information regarding the correlates of schooling attainment and that our generational approach, using three generations for a particular family, provides insights not otherwise attainable.

Using these data we examine years of schooling by birth year, gender, and rural/urban status.[7] We construct a single years of schooling measure based on 12 educational categories reported in the survey (e.g. "did not finish primary school but capable of reading and writing," "graduate from primary school", "graduate from middle school",. . .) Appendix table 3.6 reports our cross-walk between categories and years of schooling.

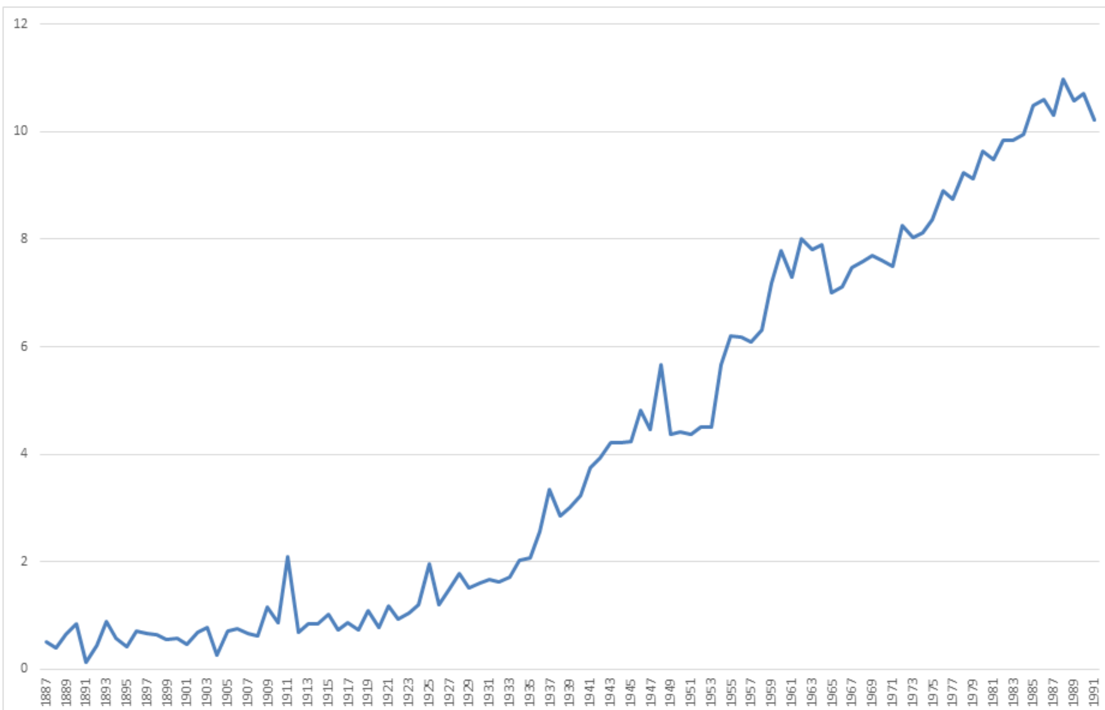


Figure 3.1: Schooling by Birth Year

Figure 3.1 shows a dramatic and continued rise in schooling levels over time for the full sample. The horizontal axis measures birth year of the individual and the vertical axis measures schooling attainment in 2013 when the youngest in our sample are 22 years old.

[7]The measure of rural/urban differs marginally across generations. For children of the respondents and the respondents themselves, we use their initial hukou. The hukou for grandparents is available only if the grandparent is still alive. If we do not have this information because the grandparent has died or because it is missing, we use, in order, whether the grandparent grew up in a rural area or currently lives in a rural area. Absent either of these two measures, we impute urban/rural status based on the region of their child's hukou (the parent's hukou in our terminology).

Figure 3.2 shows this rise in schooling by gender and clearly demonstrates the convergence in the educational attainment of males and females. Although difficult to discern in the figure, the schooling level for women has not only caught up to that of men, but has begun to surpass it. Despite changes like compulsory schooling laws and China's one child policy, we do not see dramatic breaks in trend for the 1979 or 1986 birth year cohorts in either figure. However, we do see some indication of a decline in schooling for those born in the early 1950s, who would have been approaching middle school age during the Cultural Revolution from 1966 to 1976.

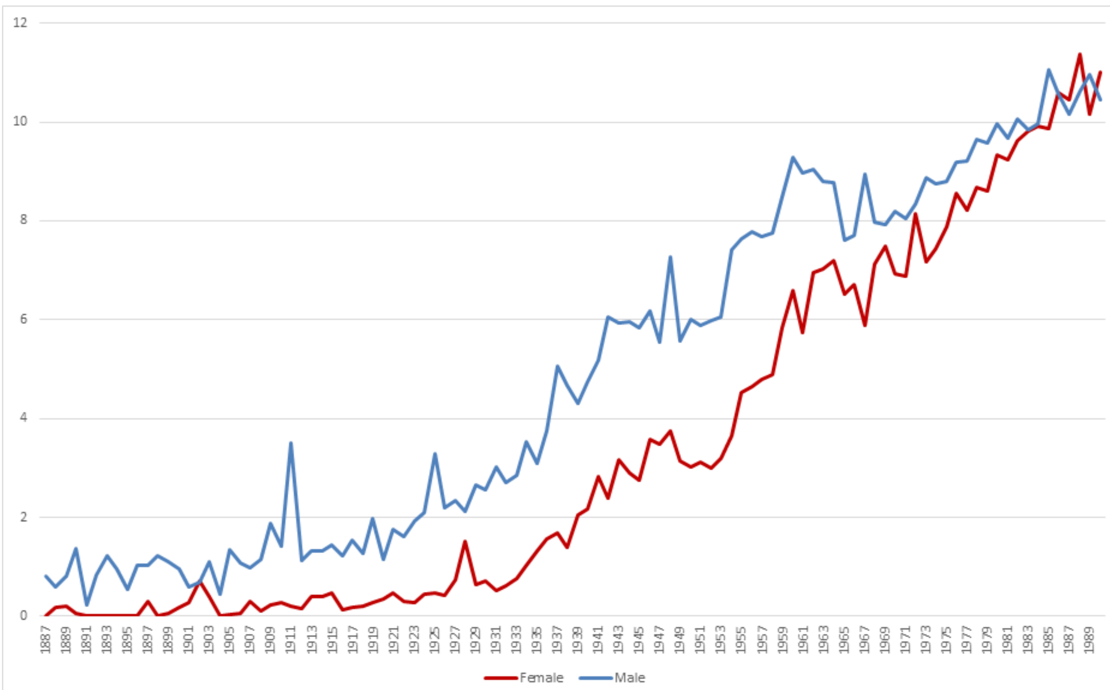


Figure 3.2: Schooling by Birth Year and Gender

Figures 3.3, 3.4, and 3.5 repeat the analysis for rural and urban individuals. As shown in figure 3.3, the rise in schooling levels was initially more rapid in urban areas (and given our smaller number of observations, more noisy) but the two regions have risen roughly in parallel for the past 30 or more years.

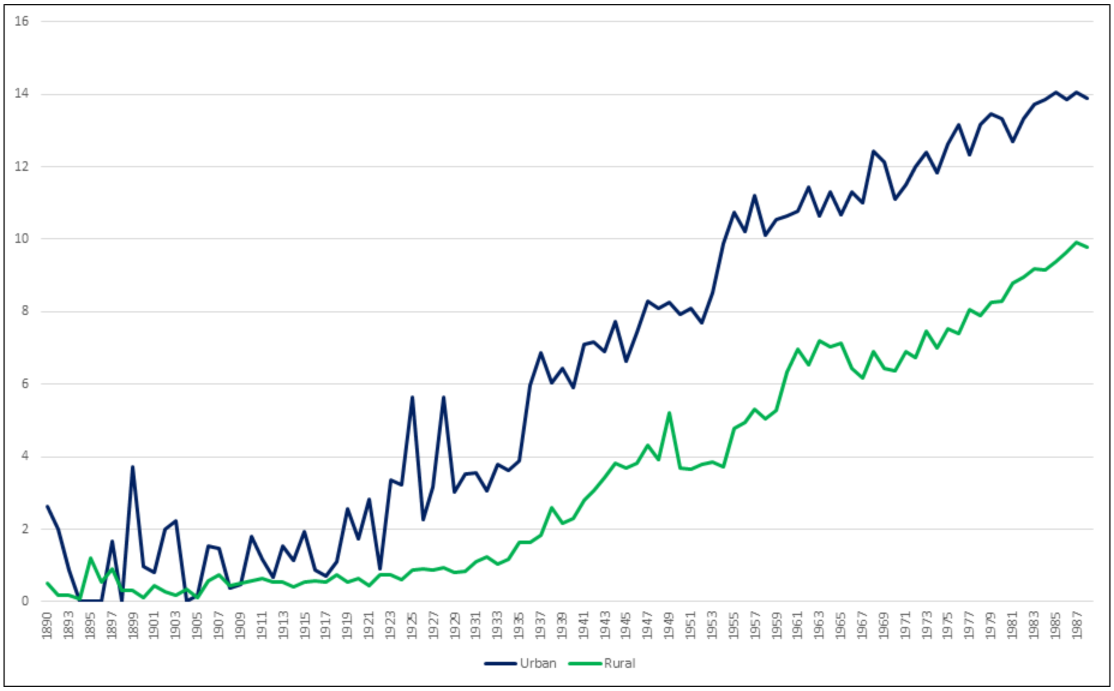


Figure 3.3: Schooling by Birth Year and Residence

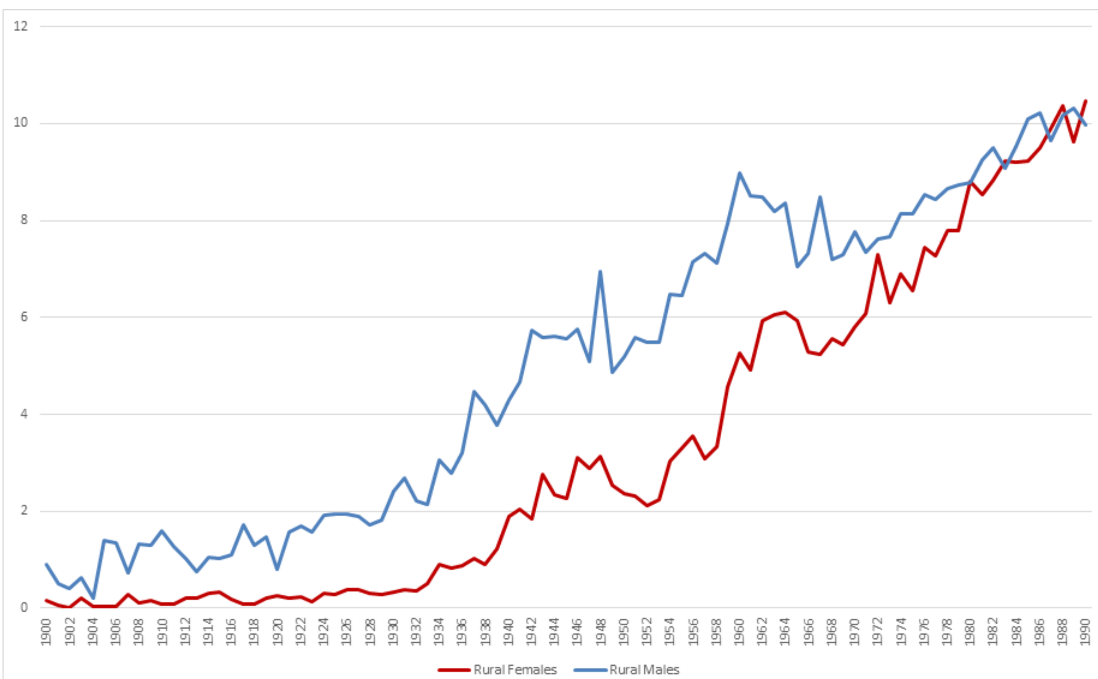


Figure 3.4: Schooling by Birth Year and Gender: Rural

Perhaps most interestingly, echoing the results in figure 3.2, the schooling attainment of women has caught up (and somewhat surpassed) that for men, even in rural areas where son preference would be thought to be most entrenched.[8] There was, however, a delayed convergence in rural areas. While women in urban areas were achieving educational levels similar to that of men for cohorts born as early as the mid-1950s, in rural areas this convergence occurred approximately 20 years later.

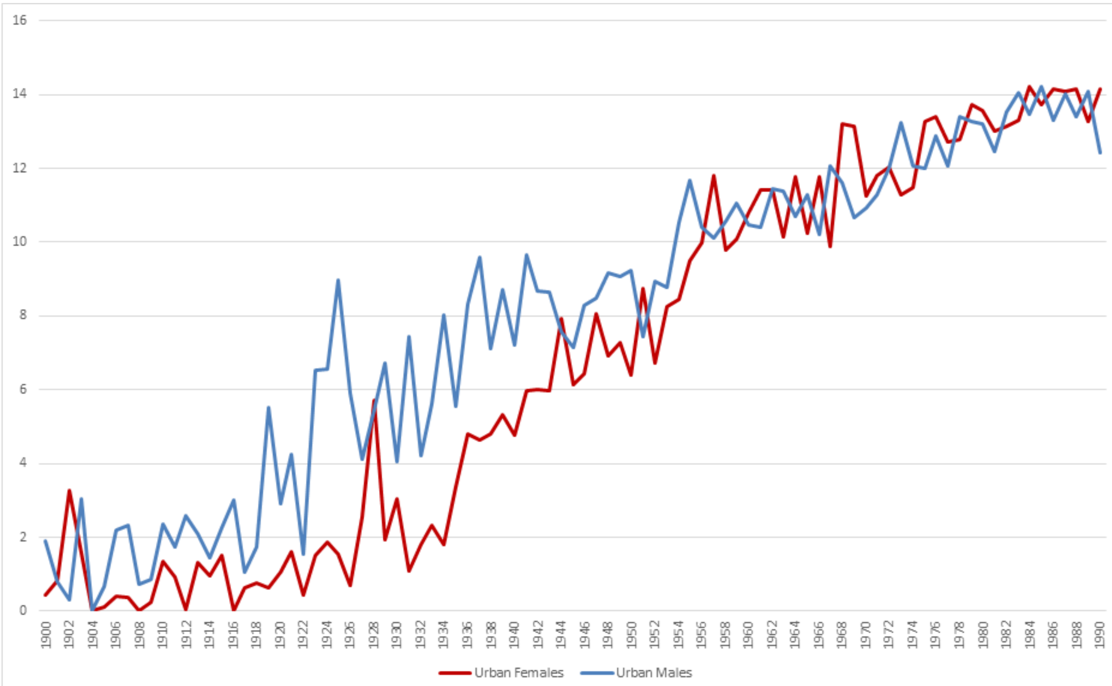


Figure 3.5: Schooling by Birth Year and Gender: Urban

To assess these increases in schooling in more detail, we examine the distribution of completed schooling. Table 3.2 shows the level of schooling attained—none (illiterate), primary school or less, middle school, high school, or college or above—for each cohort, and separately by urban and rural status. We construct these categories from survey responses that include much more detail (e.g. home schooling, literate but did not finish primary school, and type of graduate degree). Appendix table 3.6 lists the originally reported schooling level. As was apparent in the figures, the rise in educational attainment across three generations is dramatic. Average schooling for urban individuals rose from 2.83 in the oldest cohort to

[8]Recall the hypothesis of Almond, Li, and Meng (2014) that given the sex selection possible before birth and the one child policy, families with daughters may have less strong son preference, and thus greater investments in those daughters.

Table 3.2: Patterns of Education Across Cohorts

	Urban Hukou			Rural Hukou		
	Male	Female	All	Male	Female	All
Child cohort						
Number of observations	1836	1635	3471	11,419	10,374	21,793
Average birth year	1974	1974	1974	1976	1976	1976
Years of schooling	12.28	12.45	12.36	8.58	7.46	8.04
Level / degree obtained						
Illiterate (0/1)	0.44	0.87	0.65	2.89	10.09	6.35
Primary and below (0/1)	7.82	6.19	7.02	33.66	38.55	36.01
Middle School (0/1)	23.71	22.36	23.05	39.46	30.91	35.36
High School (0/1)	31.11	35.88	33.44	13.76	11.63	12.74
College and above (0/1)	36.91	34.71	35.84	10.23	8.81	9.55
Parent cohort						
Number of observations	758	777	1535	5670	6521	12,191
Average birth year	1951	1953	1952	1952	1953	1953
Years of schooling	9.40	8.32	8.85	6.32	3.39	4.84
Level / degree obtained						
Illiterate (0/1)	4.07	10.23	7.29	12.05	41.32	27.66
Primary and below (0/1)	23.84	23.28	23.55	46.78	37.80	41.99
Middle School (0/1)	29.72	28.38	29.02	26.09	15.18	20.27
High School (0/1)	32.66	29.41	30.96	12.77	5.10	8.68
College and above (0/1)	9.71	8.70	9.18	2.31	0.60	1.40
Grandparent cohort						
Number of observations	1020	1821	2841	7947	13,597	21,544
Average birth year	1923	1924	1924	1922	1924	1923
Years of schooling	4.34	1.98	2.83	1.79	0.45	0.94
Level / degree obtained						
Illiterate (0/1)	36.80	69.83	57.96	62.48	89.81	79.84
Primary and below (0/1)	39.20	19.34	26.48	30.66	8.81	16.78
Middle School (0/1)	7.99	4.89	6.00	4.20	0.95	2.14
High School (0/1)	10.55	470	6.80	2.14	0.42	1.05
College and above (0/1)	5.46	1.24	2.76	0.52	0.00	0.19

Note: Data for all three generations in a family are reported by the parent generation.

12.36 in the most recent. The comparable numbers for rural status are from less than one year of schooling to eight years.

When looking at the individual categories of schooling, it is particularly astonishing to see the high rate of illiteracy prevalent among the oldest generation. The average year of birth for grandparents is around 1923-1924 and yet nearly 60 percent of urban individuals and as many as 80 percent of rural individuals were illiterate.[9] By comparison, in the United States in the 1920s, approximately 95 percent of adults were literate.[10] (Note the United States' statistic is for the population at that time, not the cohort born at that time who would be expected to have even greater educational attainment.) Even at these extremely high rates of illiteracy, women are disadvantaged. The illiteracy rate for rural women is 90(!) percent compared to "just" 62 percent for men. And even among urban women, illiteracy is 70 percent, so urban women are less literate than even rural men. The educational achievement in China is dramatically seen at this lowest level of education in that for the child generation, illiteracy falls to well below 1 percent for urban children and to 6.35 percent rural children. Also noteworthy is the increase at the highest level of schooling with over one-third of the children in the urban sample now having at least some college education.[11]

The narrowing of the gender gap observed in the figures is even more visible in this table. Among the grandparent generation, urban men had more than twice as many years of schooling on average as did women, 4.34 versus 1.98. By the parental generation, women had almost caught up with an average of 8.32 years of schooling compared to 9.4 for men. For the most recent cohort, women had more schooling on average than men, 12.28 years for men and 12.45 years for women—a phenomenon apparent throughout the developed world.

Change in rural areas has been slower. In the grandparent generation average years of schooling for men was 4 times that of women, narrowing to just under 2 times for the parental

[9] Note that this sample is representative of parents of a representative sample of the Chinese population, they are not themselves representative.

[10] https://ourworldindata.org/literacy/ (referenced October 20, 2017)

[11] To again draw a comparison with the United States, among those 25-44, 64 percent have some college (Ryan and Bauman, 2016).

Table 3.3: Correlations in schooling across generations

	Son	Daughter	Mother	Father	Father's mother	Father's father	Mother's mother	Mother's father
Son	1.00	0.59	0.41	0.44	0.14	0.20	0.15	0.19
Daughter		1.00	0.51	0.48	0.19	0.25	0.20	0.23
Mother			1.00	0.43	0.25	0.29	0.27	0.29
Father				1.00	0.18	0.24	0.18	0.22
Father's mother					1.00	0.47	0.33	0.24
Father's father						1.00	0.27	0.33
Mother's mother							1.00	0.47
Mother's father								1.00

generation and to just 15 percent greater (8.58 versus 7.46) for the youngest cohort.

Our data are unique in that we have information on the schooling for three generations within a family. We can therefore examine the extent to which there are within family correlations across generations. Because of our interest in gender differences, we look separately at correlations for sons/daughters, mothers/fathers, and grandfathers/grandmothers. Table 3.3 presents the results. We note that there are relatively large correlations across generations that as expected, decline with the distance between the generations. The correlation between sons (daughters) and their mothers is 0.41 (0.51) and 0.44 (0.48) with fathers. It is worth noting that the correlations with both parents are slightly higher for daughters than sons. If we consider parental education to be a proxy for family income then it may suggest that income is more important in determining a daughter's educational attainment than a son's. Alternatively, greater schooling by parents may reflect a greater emphasis placed on education, an emphasis that is more important for daughters who have traditionally been short changed in this regard.

The correlations between the educational levels of children and their parents are uniformly greater than those between parents and grandparents, with the correlation between a father (mother) and his (her) own mother at 0.18 (0.27), and 0.24 (0.29) with his (her) own father. Again, the correlations are higher between women and their parents than for men and their parents. Interestingly, we also see that correlations with the educational levels of one's in-laws are nearly identical to those of own parents, perhaps attesting to assortative mating or to the similarity of educational levels within a particular locale.

Finally, with regard to within generational correlations, the correlations between married couples in the parent generation (i.e. between mothers and fathers in our nomenclature) at 0.43 is quite similar to that between married couples one generation older, 0.47 for both that between the husband's parents and the wife's parents. With respect to the children, we note the correlation between brothers and sisters at 0.59 is similar to that found in the United States and other countries for siblings more generally (not brothers and sisters). See des Etangs-Levallois and Lefranc (2017) for a summary of the literature across European countries and the United States.

3.3.3 Measurement of Key Variables

In order to assess the importance of compulsory schooling laws and the one child policy we need good measures of the extent to which the policies were relevant for a particular child. We thus construct two new variables, one for each policy, that summarize the impact. We describe our efforts here. Because of the timing of these policy interventions, they are relevant only for the schooling attainment of the most recent cohorts—those who were attending or could have been attending school in the 1970s and 1980s or whose siblings may have done so.

In 1986, China enacted a law mandating nine years of compulsory education for all children. Given the existing schooling levels, this law primarily affected rural areas wherein previous requirements were just four to six years of schooling. The law was rolled out gradually across provinces beginning in the most economically advanced areas. Children were subject to the law if they had not reached grade nine (the new required level of schooling) at the time the law went into effect. We thus code our measure of compulsory schooling to be specific to the child's birth year and province.[12] Within each province the law was introduced gradually from more to less urban areas. We do not have information at this level of detail and simply use the date at which the law first went into effect in each child's home province. In our regression analyses to follow, we interact this measure with an indicator of urban or rural status to allow for differing effects in enforcement and implementation.

The second key policy change is the advent of China's one child policy, enacted in 1979 and implemented in 1980. As with the compulsory schooling laws, enforcement varied across regions.[13] Later the government relaxed the requirement and allowed rural families to have a second child if the first child was a girl or if both parents were only children.[14]

The impact of the one child policy depends primarily on the age of the mother and on how her fertility was affected because of the law. She may have made a conscious decision not to have a daughter through selective abortion, or may have altered the timing of the pregnancy given that she would likely incur only one pregnancy. To capture the impact on the mother and thus on the child, we use a measure of exposure to family planning regimes similar to that constructed in Wang (no date) based on the fraction of a woman's fertile years spent under the one child policy.[15] In examining the schooling of children in our sample, we use the measure constructed based on the mother's age. We also interact this measure with a dummy variable for the gender of the child.

[12]The information was collected from Regulations of the Implementation of Compulsory Schooling Laws of each province.

[13]The policy also affected only the Han majority. The Han population constitutes approximately 93 percent of our sample; nearly identical to the approximately 92 percent in China as a whole. We have repeated all the analyses below with controls for ethnicity, but given the overwhelming fraction Han, our conclusions are the same regardless of whether it is included. For parsimony we report results ignoring this measure of ethnicity.

[14]Families could also have more than one child if they paid the fine associated with the additional child or if children were born outside the country.

[15]Wang used the fraction of years under each of three different family planning regimes. We use only the most recent as it is most relevant to our sample.

3.4 Data analysis

The dramatic rise in schooling shown in our figures and in table 3.2 likely has many causes. There has been widespread economic growth that has both made schooling more affordable and that has raised the returns to schooling. Also in a transition from an agricultural economy to a more urban economy, children are less needed to work the land and the opportunity cost of schooling is thus lower.[16] However, as discussed above, there have also been changes in government policies that likely influenced schooling levels, most directly the installation of compulsory schooling laws and more indirectly, the one child policy which, among other effects, would have reduced competition for parental resources, and for girls, reduced the probably that a brother would be favored.

3.4.1 Full sample

We begin with a standard regression equation for completed schooling using our stacked data for all three generations:

$$Schooling_{ij} = \alpha_0 + \alpha_1 X_{1ij} + \varepsilon_{ij}$$

Column (1) of table 3.4 reports the results for our most basic specification controlling for gender, time, and whether the person has rural hukou. We measure time in terms of birth year, centered at the birth year of the oldest individual in our sample (1852).[17] The variable thus increments by one for each year of birth beyond 1852 and for a child born in 1980, would have a value of 118. Other variables are defined straightforwardly.

[16]Compulsory schooling laws outlawed the hiring of children younger than 15 although in rural areas such employment likely continued to at least some extent.

[17]The birth years of the grandparent generation are reported by the respondents. While a grandparent born so long ago appears to be unlikely, we have not edited the data.

Table 3.4: Regression Analysis for All Three Generations

	OLS	FE	OLS	FE	OLS	FE
	(1)	(2)	(3)	(4)	(5)	(6)
Time (birth year - 1852)	0.124	0.131	0.119	0.127	0.137	0.149
	(0.001)	(.001)	(0.001)	(0.001)	(0.003)	(0.002)
Grew up in rural Area	-3.340	-1.343	-3.139	-1.350	-1.045	1.489
	(0.045)	(0.083)	(0.066)	(0.083)	(0.290)	(0.287)
Female	-1.690	-1.773	-2.365	-2.506	-6.137	-5.809
	(0.030)	(0.027)	(0.146)	(0.110)	(0.356)	(0.327)
Female*time			0.010	0.007	0.049	0.046
			(0.001)	(0.001)	(0.003)	(0.003)
Female*rural					3.907	3.747
					(0.379)	(0.347)
Rural*time					-0.020	-0.024
					(0.003)	(0.003)
Female*rural*time					-0.045	-0.043
					(0.004)	(0.004)
Constant	-3.428		-3.037		-4.893	
	(0.076)		(0.109)		(0.272)	
Observations	58,362		58,362		58,362	
Mean of Dep Variable	5.13		5.13		5.13	
R-squared	0.50	0.67	0.50	0.68	0.51	0.68

Unsurprisingly, each of these factors has a significant effect on schooling and even with this most parsimonious specification, the R2 is 0.50. Each additional year corresponds to a gain of 0.12 years of schooling, or 1.2 years per decade. We also see a large negative effect for a rural hukou, associated with a reduction of 3.3 years in expected education. Females too are worse off, with 1.7 fewer years on average.

Because we have multiple observations per family, we can control for unobserved family fixed effects. These could represent family values regarding education, resources to finance an education, or potentially measures of the difficulty of accessing schooling in the particular family's locale. Specifically, we allow the error term in the regression specification to have a family fixed effect in addition to the individual component so that the error term is $\varepsilon_{ij} + u_j$. The results of this fixed effects specification, reported in column (2), are similar to those in column (1). An additional year of time is associated with 0.13 more years of schooling, and women have 1.8 years fewer years of schooling than men. The effect of living in a rural area is mitigated when we control for family effects. This variable is identified using of families in which there is a generational shift in region.

Expanding on our list of control variables, in columns (3) and (4) we include an interaction of female with time to assess the extent to which the disadvantage women have in terms of schooling has being declining over time. With the inclusion of this interaction term, the linear effect of time is similar to that in the original specification at 1.2 years of schooling per decade, as is the coefficient on rural status. There is a slightly larger negative effect of being female as measured at time zero, but the interaction term means that this negative effect declines over time at the rate of 0.1 years per decade—an extremely modest gain for women.

The final two columns add interactions between female and rural, rural and time, and female, rural and time. These regressors allow for varying gains for women in rural relative to urban areas, different time trends for the two regions, and differing time trends for rural and urban areas more generally. As in the prior specifications, all coefficients are significantly different from zero. Time continues to have a similar effect as in column (1), with the coefficient increasing only slightly to 1.4 years of additional schooling per decade. With the addition of the interaction terms, the linear effect of a rural hukou is much larger in absolute value than it was previously, and its negative impact increases slowly over time at a rate of 0.2 years per decade. For women with a rural hukou, the negative effect of a rural hukou at time zero disappears, likely simply because the years of schooling that long ago were near zero in rural areas for all, so that there is little difference for men and women. However, the three way interaction (female*rural*time) points to a decline in the relative position of women in rural areas relative to men and relative to women in urban areas.

3.4.2 Policy Factors

To focus on our policy variables, we limit our sample to observations for the youngest cohort because they are of the age that they would have potentially been impacted by the changing policies. In table 3.5 we first replicate the results of table 3.4 for the single cohort to assess whether the standard set of regressors impact years of schooling differently for this cohort. In columns (5) through (8) we then add regressors to examine the effect of the two policy interventions on years of schooling.

In the simplest specification, column (1), the estimated effects are surprisingly similar to those shown in the first column of table 3.4. An additional year for this cohort is associated with a 0.14 gain in schooling, slightly larger but comparable to the 0.124 value in table 3.4. The effect of a rural hukou, while similar to that previously reported, is now larger in absolute value. This result stems from the rising levels of education overall; as schooling levels rise at similar rates overall, the difference between rural and urban measured in levels becomes larger. There is, however, a much larger change for the female dummy variable relative to table 3.4; for this youngest cohort, women can expect only 0.9 fewer years of schooling relative to men compared to the 1.7 years when estimated over all cohorts.

Column (3) adds the interaction of female with time. The estimated effect is much larger (nearly seven times greater) than in the parallel specification in table 3.4, pointing to the recent relative gains in female educational levels in China. As with table 3.4, the fixed effects specifications do not reveal any substantial differences.

The remaining columns in table 3.5 include measures of our policy changes, add parental schooling measures, and further interactions.[18] To examine the effect of the one child policy, we add an indicator for the child being an only child (we do not have this information for the older cohorts) as well as measures for the degree to which the child's mother was impacted

[18]We cannot include parental schooling in table 3.4 because we have no such measures for the oldest cohort.

Table 3.5: Regression Analysis for All Three Generations

	OLS	FE	OLS	FE	OLS	FE	OLS	FE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Time (birth year – 1852)	0.140	0.081	0.101	0.049	0.029	0.039	0.036	0.042
	(0.003)	(.005)	(0.004)	(0.005)	(0.005)	(0.006)	(0.006)	(0.007)
Grew up in Rural Area	-4.263	-1.386	-4.257	-1.346	-2.613	-1.385	-2.365	-0.977
	(0.076)	(0.005)	(0.076)	(0.190)	(0.079)	(0.193)	(0.124)	(0.226)
Female	-0.910	-1.226	-9.369	-9.598	-9.963	-9.695	-7.191	-8.453
	(0.052)	(0.046)	(0.678)	(0.595)	(0.638)	(0.602)	(1.069)	(1.003)
Female*Time			0.068	0.068	0.073	0.069	0.055	0.063
			(0.005)	(0.005)	(0.005)	(0.005)	(0.009)	(0.009)
Only child					1.097	–	1.273	–
					(0.121)	–	(0.148)	–
One child policy					0.645	–	0.427	–
					(0.106)	–	(0.141)	–
Compulsory schooling					0.352	0.305	-0.159	0.332
					(0.085)	(0.096)	(0.214)	(0.308)
Father's yrs of school					0.253	–	0.229	–
					(0.007)	–	(0.010)	–
Mother's yrs of school					0.222	–	0.181	–
					(0.008)	–	(0.011)	–
Female*comp school							-0.939	-0.841
							(0.320)	(0.356)
Female*one child policy							0.502	–
							(0.207)	–
Rural*comp school							0.577	-0.090
							(0.281)	(0.366)
Rural*one child policy							0.079	0.148
							(0.207)	(0.216)
Female*rural							-1.427	-1.203
							(0.183)	(0.175)
Female*rural*comp sch							0.919	1.209
							(0.418)	(0.444)
Female*rural*one child pol							0.040	
							(0.304)	
Female*rural*only child							0.387	
							(0.413)	
Female*only child							0.101	–
							(0.326)	–
Female*father's school							0.043	0.036
							(0.015)	(0.014)
Female*mother's school							0.082	0.101
							(0.016)	(0.015)
Constant	-4.709		-0.726		6.845	–	4.199	
	(0.344)		(0.468)		(0.617)	–	(0.733)	
Observations	23,473		23,473		23,473		23,473	
Mean of Dep Variable	8.54		8.54		8.54		8.54	
R-squared	0.20	0.73	0.21	0.74	0.30	0.74	0.31	0.74

by the one child policy during her child bearing years (denoted here with the label "one child policy"). In column (7) and column (8), we add variables delineating whether the child was subject to compulsory schooling laws, the educational levels of the child's father and mother, and numerous interactions with gender and rural hukou.

Beginning in column (5) we note first that the specification is well identified; all coefficients are significantly different from zero. With the addition of policy variables and parental education, the time trend decreases substantially to just 0.29 years per decade. This value pertains to males as the effect for females is the sum of this coefficient and the interaction female*time ($0.029 + 0.073 = 0.102$). Other variables have similar effects to those in the prior specifications. Among the newly added variables, being an only child is associated with over one year of additional education, a substantial increase. With a mean education level of 8.54 years, this represents a 13 percent increase and is roughly equivalent to the gain experienced by women over a decade of time.

The one child policy adds to this only child effect with an additional 0.65 years of schooling, as does the compulsory schooling law which is associated with 0.35 years of schooling. Both father's schooling and mother's schooling levels have positive and significant effects on the child's schooling attainment that are similar in magnitude. An additional year of education for the child's father is associated with 0.25 additional years for the child, while an additional year of mother's education is associated with 0.22 years. Compared to a child whose father has a middle school education, a child whose father graduated from high school would be expected to have 0.25*3 or 0.75 additional years of schooling. This effect could come directly from the value more educated parents place on schooling, the accessibility of schooling in the locale, or could serve as a proxy for financial resources of the family.[19]

When turning to a fixed effects specification (column (6)) we are unable to identify those regressors that do not vary within family. Because "family" in this one-generation specification is just siblings, several variables are constant within family, namely only child, the

[19] While we have good measures of family income, they pertain to 2011 or 2013 when the interviews were conducted, not to the time at which children were enrolled or considering enrolling in school.

one child policy measure that is based on the mother's age and location, and the years of schooling attained by the father and mother. The compulsory schooling variable is identified because children were born in different years and thus faced different regimes, although we recognize that there could be "spillover effects" if parents endeavor to ensure the same schooling for all siblings. The estimated effect of compulsory schooling is statistically indistinguishable from that in the OLS specification, as are other variables with the exception of rural hukou which again changes significantly in the fixed effects version.

In the final pair of columns we add interaction terms allowing compulsory schooling laws, the one child policy, and parental education to vary across urban and rural areas and by the gender of the child. With the large number of interactions, the net effects are often difficult to discern with a passing glance. Here we focus on relative rather than absolute comparisons.

Focusing on these linear and interaction terms, being an only child continues to have a large, positive and significant effect on schooling, as does the one child policy. However, the linear term for compulsory schooling is not significantly different from zero and in fact, the standard error is larger than the coefficient. Regarding interactions with "female," being the only child has a significantly larger effect for daughters, suggesting that competition and tighter financial constraints matter more for the schooling of girls. Perhaps surprisingly, the one child policy does not appear to favor girls over boys; the linear effect is positive and significant but the interaction with female is not significantly different from zero. (The point estimate is 0.50 with a standard error of 0.2.) Schooling of parents continues to be more important, but echoing the results in table 3, the relationship is significantly stronger for daughters than sons. An additional year of education for one's father is associated with 0.23 years of additional schooling for a son and $0.23 + 0.043 = 0.273$ for a daughter. If schooling is a proxy for the family's financial means, this result would indicate the income elasticity of schooling is greater for daughters than sons. Interestingly, an additional year of schooling for a mother implies 0.18 years of additional schooling for a son, but $0.18 + 0.082 = 0.262$ for a daughter. Thus, while the effect of either parent's education is stronger for daughters than sons, the additional effect of a mother's education for daughters is greater than the additional effect of a father's perhaps pointing to the greater bargaining position of more educated wives.

As expected, children in rural areas are worse off, and the penalty is larger for girls. While the linear term for compulsory schooling is itself not significantly different from zero, compulsory schooling does have a large effect in rural areas of just over one-half of a year of schooling (significantly different from zero at the 5 percent level). And this positive effect is substantially larger still for girls in rural areas, associated with an additional year of education.

When looking at the fixed effects specification, the results are all similar to the OLS specification with the exception of the effect of the mother's schooling for daughters which increases while that for father's schooling decreases.

As these results demonstrate, public policy can have important effects on schooling attainment as well as on male/female and rural/urban differences in this measure. We note that the gains in education observed here come not just from policies that directly target education, such as compulsory schooling laws, but from policies such as the one child policy that have indirect effects. Depending on the specification, the one child policy and compulsory schooling have effects that are far larger than simply economic growth as proxied by a time trend.

3.5 Conclusion

The rise in economic growth and educational attainment experienced in China over the last several decades has been stunning and have been shared across demographic groups improving outcomes for both men and women and those in rural and urban areas. Further, the gains experienced, particularly by women, resulted in a reduction in the large disparity initially existing between schooling levels of men and women. In fact, we find that women now have more average years of schooling than do men. Gains in rural areas, while substantial still leave individuals in these areas far behind their urban counterparts.

In this paper we examine the effect on schooling of two important policy changes—the institution of nine years compulsory schooling and the one child policy. Our study differs from past work in that we can examine differences within families and across generations of the same family. We find that the effects of compulsory schooling were large and significant only in rural areas and were largest for rural women. This result is not surprising as schooling in urban areas was already typically higher than that mandated by the policy. In contrast, the one child policy, which was instituted nationwide but which was more strictly enforced in urban areas, had nearly equal positive effects on schooling for men and women and for rural and urban areas. These results point more generally to the importance of policy as well as general economic growth in driving schooling attainment.

Appendix

Table 3.6: Assigned years of schooling based on education level

Education level	Assigned Years of Schooling
No formal education (illiterate)	0
Did not finish primary school but capable of reading or writing	2
Sishu/home school	2
Graduate from elementary school	6
Graduate from middle school	9
Graduate from high school	12
Graduate from vocational school	14.5
Graduate from Two/Three Year College/Associate degree	14.5
Graduate from Four Year College/Bachelor's degree	16
Graduate from Post-graduate, Master's degree	18
Graduate from Post-graduate, Doctoral degree/Ph.D.	21

Bibliography

Abdulkadiroğlu, Atila, Nikhil Agarwal, and Parag A. Pathak. 2017. "The Welfare Effects of Coordinated Assignment: Evidence from the New York City High School Match." *American Economic Review* 107 (12):3635–3689.

Abdulkadiroğlu, Atila and Tayfun Sönmez. 2003. "School Choice: A Mechanism Design Approach." *American Economic Review* 93 (3):729–747.

Agarwal, Nikhil. 2015. "An Empirical Model of the Medical Match." *American Economic Review* 105 (7):1939–1978.

Agarwal, Nikhil and Paulo Somaini. 2018. "Demand Analysis Using Strategic Reports: An Application to a School Choice Mechanism." *Econometrica* 86 (2):391–444.

Almond, Douglas, Hongbin Li, and Lingsheng Meng. 2014. "Son Preference and Early Childhood Investments in China."

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

Berry, Steven T. and Philip A. Haile. 2009. "Nonparametric identification of multinomial choice demand models with heterogeneous consumers." Tech. rep., National Bureau of Economic Research.

Brewer, Dominic J., Eric R. Eide, and Ronald G. Ehrenberg. 1999. "Does It Pay to Attend an

Elite Private College? Cross-Cohort Evidence on the Effects of College Type on Earnings." *The Journal of Human Resources* 34 (1):104–123.

Briesch, Richard A., Pradeep K. Chintagunta, and Rosa L. Matzkin. 2010. "Nonparametric Discrete Choice Models With Unobserved Heterogeneity." *Journal of Business & Economic Statistics* 28 (2):291–307.

Bucarey, Alonso. 2018. "Who Pays for Free College? Crowding Out on Campus." *Working Paper, MIT Department of Economics* .

Che, Yeon-Koo, Jinwoo Kim, and Fuhito Kojima. 2019. "Stable Matching in Large Economies." *Econometrica* 87 (1):65–110.

Che, Yeon-Koo and Youngwoo Koh. 2016. "Decentralized College Admissions." *Journal of Political Economy* 124 (5):1295–1338.

Chiappori, Pierre-André and Ivana Komunjer. 2009. "On the nonparametric identification of multiple choice models." *Manuscript, Columbia University* .

Chiappori, Pierre-André and Bernard Salanié. 2016. "The Econometrics of Matching Models." *Journal of Economic Literature* 54 (3):832–861.

Choo, Eugene and Aloysius Siow. 2006. "Who Marries Whom and Why." *Journal of Political Economy* 114 (1):175–201.

Chyi, Hau and Bo Zhou. 2014. "The effects of tuition reforms on school enrollment in rural China." *Economics of Education Review* 38:104–123.

Connelly, Rachel and Zhenzhen Zheng. 2003. "Determinants of school enrollment and completion of 10 to 18 year olds in China." *Economics of Education Review* 22 (4):379–388.

Cosslett, Stephen R. 1983. "Distribution-Free Maximum Likelihood Estimator of the Binary Choice Model." *Econometrica* 51 (3):765–782.

Dale, Stacy Berg and Alan B. Krueger. 2002. “Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables.” *The Quarterly Journal of Economics* 117 (4):1491–1527.

Das Gupta, Monica, Zhenghua Jiang, Bohua Li, Zhenming Xie, Chung Woojin, and Hwa-Ok Bae. 2003. “Why is Son preference so persistent in East and South Asia? a cross-country study of China, India and the Republic of Korea.” *The Journal of Development Studies* 40 (2):153–187.

des Etangs-Levallois, Céline Lecavelier and Arnaud Lefranc. 2017. “Sibling correlations in terms of education, profession and earnings, in France.” Tech. Rep. 2017-12, THEMA (THéorie Economique, Modélisation et Applications), Université de Cergy-Pontoise.

Diamond, William and Nikhil Agarwal. 2017. “Latent indices in assortative matching models.” *Quantitative Economics* 8 (3):685–728.

Eliaz, Kfir and Ran Spiegler. 2011a. “Consideration Sets and Competitive Marketing.” *The Review of Economic Studies* 78 (1):235–262.

———. 2011b. “On the strategic use of attention grabbers.” *Theoretical Economics* 6 (1):127–155.

Epple, Dennis, Richard Romano, Sinan Sarpça, and Holger Sieg. 2017. “A general equilibrium analysis of state and private colleges and access to higher education in the U.S.” *Journal of Public Economics* 155:164–178.

Epple, Dennis, Richard Romano, and Holger Sieg. 2006. “Admission, Tuition, and Financial Aid Policies in the Market for Higher Education.” *Econometrica* 74 (4):885–928.

Fack, Gabrielle, Julien Grenet, and Yinghua He. 2019. “Beyond Truth-Telling: Preference Estimation with Centralized School Choice and College Admissions.” *American Economic Review* 109 (4):1486–1529.

Fox, Jeremy T. 2010. “Identification in matching games.” *Quantitative Economics* 1 (2):203–254.

———. 2018. “Estimating matching games with transfers.” *Quantitative Economics* 9 (1):1–38.

Fox, Jeremy T., Chenyu Yang, and David H. Hsu. 2018. “Unobserved Heterogeneity in Matching Games.” *Journal of Political Economy* .

Fu, Chao. 2014. “Equilibrium Tuition, Applications, Admissions, and Enrollment in the College Market.” *journal of political economy* :57.

Gale, David and Lloyd Shapley. 1962. “College Admissions and the Stability of Marriage.” *The American Mathematical Monthly* 69 (1):9–15.

Galichon, Alfred and Yu-Wei Hsieh. 2018. “A Model of Decentralized Matching Without Transfers.” *Working Paper* .

Galichon, Alfred and Bernard Salanie. 2015. “Cupid’s Invisible Hand: Social Surplus and Identification in Matching Models.” *Working Paper* .

Graham, Bryan S. 2013. “Comparative Static and Computational Methods for an Empirical One-to-one Transferable Utility Matching Model.” In *Advances in Econometrics*, vol. 31, edited by Eugene Choo and Matthew Shum. Emerald Group Publishing Limited, 153–181.

Härdle, Wolfgang and A. B. Tsybakov. 1993. “How sensitive are average derivatives?” *Journal of Econometrics* 58 (1):31–48.

Hemphill, C, K Nauer, H Zelon, and T Jacobs. 2009. “The new marketplace: How small-school reforms and school choice have reshaped New York City’s high schools.” *New School Center for New York City Affairs* .

Hesketh, Therese and Zhu Wei Xing. 2006. “Abnormal sex ratios in human populations: Causes and consequences.” *Proceedings of the National Academy of Sciences* 103 (36):13271–13275.

Horowitz, Joel L. 1992. “A Smoothed Maximum Score Estimator for the Binary Response Model.” *Econometrica* 60 (3):505–531.

Horowitz, Joel L. and Wolfgang Härdle. 1996. “Direct Semiparametric Estimation of Single-Index Models with Discrete Covariates.” *Journal of the American Statistical Association* 91 (436):1632–1640.

Hoxby, Caroline and Christopher Avery. 2013. “The Missing "One-Offs": The Hidden Supply of High-Achieving, Low-Income Students.” *Brookings Papers on Economic Activity* 2013 (1):1–65. URL `http://muse.jhu.edu/content/crossref/journals/brookings_papers_on_economic_activity/v2013/2013.1.hoxby.html`.

Ichimura, Hidehiko. 1993. “Semiparametric least squares (SLS) and weighted SLS estimation of single-index models.” *Journal of Econometrics* 58 (1):71–120.

Kingdon, G. Gandhi. 2002. “The Gender Gap in Educational Attainment in India: How Much Can Be Explained?” *The Journal of Development Studies* 39 (2):25–53.

Klein, Roger W. and Richard H. Spady. 1993. “An Efficient Semiparametric Estimator for Binary Response Models.” *Econometrica* 61 (2):387–421.

Lei, Xiaoyan, Kathleen McGarry, Xiaoting Sun, and Yaohui Zhao. 2017. “Gender Differences in Parental Investments and Intergenerational Transfers in China.” *Working Paper* .

Lewbel, Arthur. 2000. “Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables.” *Journal of Econometrics* 97 (1):145–177.

Li, Li and Haoming Liu. 2014. “Primary school availability and middle school education in rural China.” *Labour Economics* 28:24–40.

Li, Shuzhuo, Chuzhu Zhu, and Marcus W. Feldman. 2004. "Gender differences in child survival in contemporary rural China: a county study." *Journal of Biosocial Science* 36 (1):83–109.

Liu, Tracy Xiao, Zhixi Wan, and Chenyu Yang. 2018. "The Efficiency of A Dynamic Decentralized Two-sided Matching Market." *Working Paper* .

Manski, Charles F. 1975. "Maximum score estimation of the stochastic utility model of choice." *Journal of Econometrics* 3 (3):205–228.

———. 1985. "Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator." *Journal of Econometrics* 27 (3):313–333.

Manzini, Paola and Marco Mariotti. 2014. "Stochastic Choice and Consideration Sets." *Econometrica* 82 (3):1153–1176.

Masatlioglu, Yusufcan, Daisuke Nakajima, and Erkut Y. Ozbay. 2012. "Revealed Attention." *American Economic Review* 102 (5):2183–2205.

Matzkin, Rosa L. 1991. "Semiparametric Estimation of Monotone and Concave Utility Functions for Polychotomous Choice Models." *Econometrica* 59 (5):1315–1327.

———. 1992. "Nonparametric and Distribution-Free Estimation of the Binary Threshold Crossing and The Binary Choice Models." *Econometrica* 60 (2):239–270.

———. 1993. "Nonparametric identification and estimation of polychotomous choice models." *Journal of Econometrics* 58 (1):137–168.

———. 2007. "Heterogeneous choice." *Econometric Society Monographs* 43:75.

———. 2012. "Identification in nonparametric limited dependent variable models with simultaneity and unobserved heterogeneity." *Journal of Econometrics* 166 (1):106–115.

———. 2015. “Estimation of nonparametric models with simultaneity.” *Econometrica* 83 (1):1–66.

———. 2018. “Constructive identification in some nonseparable discrete choice models.” *Journal of Econometrics* .

McFadden, D. 1973. “Conditional logit analysis of qualitative choice behavior.” *Frontiers in Econometrics* :105–142.

———. 1981. “Econometric Models of Probabilistic Choice.” *Structure Analysis of Discrete Data with Econometric Applications* .

McGarry, Kathleen and Xiaoting Sun. 2018. “Three generations of changing gender patterns of schooling in China.” *Journal of the Asia Pacific Economy* 23 (4):584–605.

Menzel, Konrad. 2015. “Large Matching Markets as Two-Sided Demand Systems.” *Econometrica* 83 (3):897–941.

Mindruta, Denisa, Mahka Moeen, and Rajshree Agarwal. 2016. “A two-sided matching approach for partner selection and assessing complementarities in partners’ attributes in inter-firm alliances.” *Strategic Management Journal* 37 (1):206–231.

Mishra, Vinod, T. K. Roy, and Robert D. Retherford. 2004. “Sex Differentials in Childhood Feeding, Health Care, and Nutritional Status in India.” *Population and Development Review* 30 (2):269–295.

Newey, Whitney K. 1994. “Kernel Estimation of Partial Means and a General Variance Estimator.” *Econometric Theory* 10 (2):233–253.

Newey, Whitney K. and Daniel McFadden. 1994. “Large Sample Estimation and Hypothesis Testing.” In *Handbook of Econometrics*, vol. 4. Elsevier, 2111–2245. URL `http://www.sciencedirect.com/science/article/pii/S1573441205800054`.

Pande, Rohini P. 2003. "Selective gender differences in childhood nutrition and immunization in rural India: The role of siblings." *Demography* 40 (3):395–418.

Powell, James L., James H. Stock, and Thomas M. Stoker. 1989. "Semiparametric Estimation of Index Coefficients." *Econometrica* 57 (6):1403–1430.

Ryan, Camille L. and Kurt Bauman. 2016. "Educational Attainment in the United States: 2015." Tech. rep. URL `https://www.census.gov/content/dam/Census/library/publications/2016/demo/p20-578.pdf`.

Sinha, Shruti. 2015. "Identification and Estimation in One-to-One Matching Models with Nonparametric Unobservables." *Working Paper, Toulouse School of Economics* .

Song, Lina, Simon Appleton, and John Knight. 2006. "Why Do Girls in Rural China Have Lower School Enrollment?" *World Development* 34 (9):1639–1653.

Terry Long, Bridget. 2004. "How have college decisions changed over time? An application of the conditional logistic choice model." *Journal of Econometrics* 121 (1-2):271–296.

The World Factbook. 2018. URL `https://www.cia.gov/library/publications/the-world-factbook/fields/2018.html`. Accessed 19 March 2018.

Wang, Fei. no date. "Using New Measure to Reassess the Fertility Effects of China's Family Planning Policies."

Wang, Wendy. 2005. "Son preference and educational opportunities of children in China—"I wish you were a boy!"." *Gender Issues* 22 (2):3–30.

World Bank. 2018. "Sex ratio at birth (male births per female births) | Data." URL `https://data.worldbank.org/indicator/SP.POP.BRTH.MF?locations=CN&view=chart`. Accessed 5 September 2018.