Essays in Empirical Matching

Essays in Empirical Matching

A dissertation presented
by

Nikhil Agarwal

to
The Committee on Degrees in Economics
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of

Economics

Harvard University
Cambridge, Massachusetts
May 2013

Thesis Advisor Author

Ariel Pakes **Nikhil Agarwal**

Essays in Empirical Matching

Abstract

This thesis combines three essays on empirical applications and methods in two-sided matching markets. The first essay uses existing methods to estimate preferences for schools using rank order lists from New York City's new high school assignment system launched in Fall 2003 to study the consequences of coordinating school admissions in a mechanism based on the student-proposing deferred acceptance algorithm. The second essay develops techniques for estimating preferences in two-sided matching markets with non-transferable utility using only data on final matches. It uses these techniques to estimate preferences in the market for family medicine residents. These estimates are then used to analyze two economic questions. First, it investigates whether centralization in the market for medical residents is primarily responsible for low salaries paid to medical residents. Second, it analyzes the effects of government interventions intended to encourage training of medical residents in rural areas. The final essay studies estimation and non-parametric identfication of preferences in two-sided matching markets with non-transferable utility. It studies the special case in which preferences of each side of the market is vertical and data from a pairwise stable match, in a single large market is observed.

Acknowledgments

I am indebted to my advisors Susan Athey, Ariel Pakes, Parag Pathak and Al Roth for their unwavering support and guidance. Each of my advisors have uniquely influenced my work. I am particularly grateful to my co-authors Atila Abdulkadiroglu, Parag Pathak, and William Diamond who have contributed immensely to chapters one and three of this dissertation. I thank Rebecca Diamond, Dr. Joel Katz and Greg Lewis for several helpful discussions and for their generosity with their time. I also thank several members of the Harvard community for suggestions that have greatly improved this work. I dedicate this dissertation to my family.

Contents

Introduction

In many markets, two distinct set of agents match with each other based on their preferences and highly individualized prices are often not used to clear markets. Thus, agents cannot choose to match with their most preferred mate since they must be chosen as well. Examples of such marketplaces include the marriage markets, high-school or college admissions as well as some labor markets. A large body of work has been dedicated to the theoretical study of such markets (Roth and Sotomayor, 1992). This influential work has been used in the last two decades to guide the design of real-world matching markets. However, to answer some economic questions, particularly quantitatively, it is important to understand the preferences of agents participating in these markets since these primitives determine the final outcomes. Questions of interest may include the economic effects of the organization and design of such market-places as well as the impact and role of policy interventions in these markets.

This dissertation contains three essays on empirical methods and applications in two-sided matching markets with non-transferable utility. Chapters 1 and 2 are primarily empirical in focus, although Chapter 2 lays out a methodological framework for analyzing these markets. Chapter 3 studies econometric questions related to identification and estimation for the type of data considered in Chapter 2 for a simpler preference model. The essays treat the preferences of agents as the primitives of interest for analyzing the economic questions described earlier.

A significant barrier to the empirical study of preferences in these markets is the rarity of data speaking directly to these primitives. Except in rare cases when survey data or a centralized marketplace which provides truthfully reported rank-order lists, we may not be able to observe these primitives. Chapter 1 demonstrates how techniques developed for demand estimation (Beggs et al., 1981; Berry et al., 1995) can be used to analyze rank-ordered preference data.. However, such data are often not collected or, as in the case of the National Residency Matching Program, are considered highly confidential. Chapter 2 develops an emprical framework for estimating preferences using data only on the final matches, expanding our ability to analyze the primitives in these markets. The main assumptions in this framework are that data from a pairwise stable match is observed and that preferences on one side of the market are homogeneous. Methodologically, the assumptions guarantee the uniqueness of a pairwise stable match and, hence, a computationally tractable simulation algorithm. Even with the restrictive assumption on preferences on one side of the market, the chapter opens many new methodological questions about the use of data from these markets. Particularly, unlike single-agent choices, matches depend on the preferences of other agents in the market. This interdependence opens novel econometric problems on identification and estimation. Chapter 3 presents the first theoretical results on the properties of these estimators for the case when preferences on both sides of the market are vertical. The results on identification show that observing data from a market with many-to-one matching market is particularly important from an empirical perspective. While preferences are not identified with data from one-to-one matching, the model is non-parametrically identified with data from many-to-one matching. This difference in identifiability of the model can be illustrated using simulated objective functions that do and do not use information available in many-to-one matching. The essay proposes an estimator that uses both these types on information, and proves consistency of a method of moments based estimator as the size of the matching market increases, but data only on one market is observed. This consistency result does not follow directly from previous arguments because of the interdependence in matches within a single market. Finally, the chapter presents Monte Carlo studies of a simulation based estimator.

Estimates of preferences can be an important input into analyses of retrospective or proposed policy interventions as well as understanding properties of equilibria under counterfactual market structures. Such empirical investigations are central motivations in Chapters 2 and 3.

Chapter 1 uses preference estimates to study the welfare effects of the 2003 redesign of New York City's high school admissions process. Since the equilbria under the uncoordinated mechanism is not well understood from a theoretical perspective, a retrospective case study is the only available approach to understand the consequences of adopting a mechanism based on the Gale-Shapley algorithm. Compared to the prior mechanism with multiple offers and a limited number of choices, there is a 40% increase in enrollment at assigned school. The old mechanism restricted choices and placed many students close to home, while the new mechanism assigns the typical students to schools 0.79 miles further from home. Since student preferences trade off proximity and school quality, but are substantially heterogeneous, the chapter uses a distance metric welfare criterion to show the overall and the distributive effects of the reform. Even though students prefer closer schools, the new mechanism is more likely to assign students to schools they prefer and this more than compensates for the distance increase. The typical student's welfare increases by the equivalent of 0.24 miles from the new mechanism. Students from across most demographic groups, boroughs, and baseline achievement categories obtain a more preferred assignment on average from the new mechanism, suggesting that allocative changes involving assignment mechanisms need not be zero-sum. The chapter also quantifies the welfare costs of constraints such as strategy-proofness and stability placed on the mechanism by the Gale-Shapley algorithm.

Chapter 2 uses preference estimates to address two important policy issues concerning the market for medical residents. First, using both theory and empirical evidence, it examines the anti-trust allegation that the clearinghouse restrains competition, resulting in salaries below the marginal product of labor. Counterfactual simulations of a competitive wage equilibrium show that residents' willingness to pay for desirable programs results in estimated salary markdowns ranging from \$23,000 to \$43,000 below the marginal profuct of labor, with larger markdowns at more desirable programs. Therefore, a limited number of positions at high quality programs, not the design of the match, is the likely cause of low salaries. Second, I analyze wage an supply policies aimed at increasing the number of residents training in rural areas while accounting for general equilibrium effects from the matching market. The main findings are that financial incentives increase the quality, but not the number of rural residents. Quantity regulations, on the other hand, increase the number of rural trainees, but the impact on resident quality depends on the design of the intervention.

The empirical applications studied in these essays show that understanding the preferences of agents participating in matching markets allows analyzing interesting questions about these markets. In the absense of personalized prices, these primitives directly affect the assignment of heterogeneous agents. The econometric methodology developed in the final essay is a first-step towards a toolbox for leveraging the most commonly available data on such markets.

Chapter 1

Sorting and Welfare Consequences of Coordinated School Admissions: Evidence from New York City

1.1 Introduction

Public school choice has become an increasingly common part of the educational landscape. Whether choice involves transferring to an out-of-zone public school, applying to a public charter, magnet or selective exam school, or using a voucher to attend a private school, a growing number of families now have the option to opt out of their neighborhood school. As choice options have proliferated, so too has the way in which choices are expressed and students are assigned. Ad hoc decentralized assignment systems where students applied separately to schools without coordinated admissions have been replaced with centralized, coordinated application mechanisms in a number of cities including Denver, London, New Orleans and New York City. In these cities, student demand is matched with the supply of school seats using algorithms inspired by the theory of matching markets (Gale and Shapley, 1962; Shapley and Scarf, 1974). Changes in assignment protocols are inevitably controversial since reallocating school seats can result in some students ending up with high quality schools leaving others with less desirable options.

In this paper, we exploit a large-scale policy change in New York City's public schools to study the impacts of moving from an uncoordinated assignment system run by the district to a coordinated single-offer choice system on the allocation of students to schools. Prior to 2003, roughly 90,000 aspiring high school students applied to five out of more than 600 school programs; they could receive multiple offers and be placed on wait lists. Students in turn could accept only a single school offer and a single wait list offer, and the cycle of offers and acceptances repeated two more times. In many cases, top-performing students were admitted to all of their choices (Herszenhorn, 2004). However, a total of three rounds of processing applications were insufficient to allocate all students to schools, and more than 30,000 students were assigned to a school not on their original application list through an administrative process. Even though students initially expressed their preferences on a common application, admissions were not coordinated across schools and therefore we refer to the old mechanism as decentralized. Many believed that this system favored students from strong academic backgrounds, who were sophisticated enough to navigate the process. Moreover, influential parents could lobby for and obtain places for their children since some principals were able to enroll students by sidestepping the central enrollment office (Hemphill and Nauer, 2009).

In Fall 2003, the system was replaced by a single-offer choice system, with a main round based on the student-proposing deferred acceptance algorithm, where participants could rank up to 12 programs. A supplementary round assigned those unassigned in the main round. This new mechanism would place students into high schools for the 2004-05 school year. The rationale for the reform was to provide a greater voice to student's choices, to utilize school places more efficiently, and to reduce 'gaming' involved to obtain a school seat (Kerr, 2003). Parents were also provided with additional information on school options through an expanded set of parent workshops and high school fairs. At the conclusion of the first year, only about 8,000 students participated in the supplementary round, and a total of 3,000 students were administratively assigned.

This large-scale policy change provides a unique opportunity to investigate the distributional consequences of centralized and coordinated school assignment. Changes were advertised widely in New York beginning only in September 2003, and students submitted their preferences in November 2003 limiting the scope for participants to react by moving across New York. Moreover, rich micro-level data from both systems allow us to surmount many of the usual challenges associated with studying decentralized mechanisms and evaluating assignment mechanisms. For instance, in Niederle and Roth (2003b)'s pioneering study of the new centralized matching mechanism for the gastroentrology labor market, data limitations require focusing on overall mobility rates, without quantifying its tradeoffs with training program quality. A key strength of our study is that we observe the final placement of all students in the decentralized mechanism in 2002-03 as well as subsequent school enrollment, while in the new mechanism we observe the submitted rank order lists, assignments, and subsequent enrollment. Using the submitted preferences together with detailed information on student and school attributes, we can estimate what factors influence the demand for schools. The estimated preference distribution can then be used to evaluate the distributional consequences of the new mechanism and to quantitatively assess mechanism design choices.

The new mechanism is based on the student-proposing deferred acceptance mechanism, where truth-telling is a weakly dominant strategy for participants (Dubins and Freedman, 1981; Roth, 1982). This motivates our empirical strategy of estimating the preference distribution using the revealed preferences of students with a discrete choice model, where the richness of our data allows us to incorporate substantial student heterogeneity and unobserved school effects to describe student preferences. There are some complications with this approach that we thoroughly investigate, however. First, the actual mechanism in New York City only allows participants to rank at most 12 choices. This constraint interferes with the dominant strategy property of the mechanism for students who find more than 12 schools acceptable (Abdulkadiroğlu et al., 2009; Haeringer and Klijn, 2009). The advice given to participants when submitting their rank ordering is: "You must now rank your 12 choices according to your true preferences" (NYC Department of Education 2003). In our sample, 77% of applicants rank fewer than 12 schools, so their incentives are not impacted by this constraint. Second, it is possible that the abruptness of the new mechanism led some families to misinterpret the advice provided by the Department of Education, and to submit preferences according to some heuristics learned from the old mechanism. For instance, some families may behave strategically by ranking a safety school as their last choice, even though it is unnecessary had they ranked fewer than 12 choices. Directly modeling strategic choices is complicated by the fact that equilibrium reports may be influenced by small changes in information, which we do not observe. However, we anticipate these effects to be more pronounced in highly manipulable mechanisms, where strategic concerns could dominate the decision of which schools are most preferred (Hastings and Weinstein, 2008).

We begin with the assumption that all choices are truthful, but then examine variations from this benchmark. We report estimates from four variations: treating only top choice as the most preferred alternative, treating only the top three choices as the three most preferred, restricting to the sample of students who rank less than 12 schools and treating these rankings as truthful, and only using choices other than the last choice to eliminate a possible "safety-school" effect in the last choice. The estimates under each assumption are quantitatively similar to those from our benchmark model. We also report estimates in which the stated choices are not assumed to be preferred to unranked schools, and from models with different assumptions on the outside option. The estimates only using choices among ranked alternatives differs more substantially and the model fit is worse. We find the behavioral assumptions for both approaches less appealing, as we describe below.

Changes in school placement due to coordination of admissions can happen through many channels. First, the new mechanism allows students to rank seven more schools than in the old mechanism. Second, the limited number of rounds of offers and acceptances can lead to coordination failures where students hold on to less preferred choices waiting to be offered seats at more preferred schools once others decline. With limited time for market clearing, this may, in turn, lead to inefficiencies, leaving many students unassigned.[1] When many students are unassigned, the district simply placed students to the school closest to their home, even though students may have wanted to go elsewhere. In a centralized single-offer system, school seats will be used effectively, as the computerized rounds of offers and acceptances minimize coordination issues related to insufficient time. Third, in New York's old mechanism, schools were able to see the entire rank ordering of applicants, and often advertised they would only consider those who ranked them first. This property creates strategic pressure on the student's ranking decisions. Each of these features of decentralized assignment can result in mismatch of students to their preferred schools.

---

[1]This feature has been called congestion by Roth and Xing (1997) and has been explored experimentally in Kagel and Roth (2000).

On the other hand, centralized choice may make more students likely to take advantage of choice options and encourage movement throughout the district. Ravitch (2011), for instance, argues that the elevation of choice in New York City "destroyed the concept of neighborhood schools" as "children scattered across the city in response to the lure of new, unknown small schools with catchy names, or were assigned to schools far from home." Similarly, critics of the new system believed that denying principals' information on students' ranking of the school limits a principals ability to attract "students who want them most" (Herszenhorn, 2004). This logic implies that removing a school's ability to know whether they were ranked first could lead to a drop in enrollment at the assigned school. It also may be in a district's interest to advantages certain student subgroups to prevent them from exiting (Engberg et al., 2010). Providing students with multiple offers may be a way to attract those who have good outside options.[2] Taken together, these points suggest that the new mechanism may involve a substantial redistribution among students, rather than better assignments for most students.

The questions we examine in New York City are also relevant to other settings where uncoordinated assignment processes have been replaced with coordinated ones. For instance, in England, where the 2003 Admissions Code mandated coordination of admissions nationwide, a number of local education authorities, governing bodies similar to U.S. school districts, adopted common applications and centralized assignment. The motivation for the policy change exhibits striking parallels with those in New York. Authorities wanted to ensure that "every child within a local authority area would receive one offer of a school place on the same day. This would eliminate or largely eliminate multiple offers and free up places for parents who would not otherwise be offered a place" (Pennell et al., 2006). All 32 London boroughs coordinated to establish a Pan London Admissions Scheme intended to make the admissions system "fairer" and "simpler," and to "result in more parents getting an offer of a place for their child at one of their preferred schools earlier and fewer getting no offer at all" (Association of London Government 2005). Interestingly, the

[2] Williams (2004) profiles families who had to pay a deposit to reserve a spot in a private or parochial school while awaiting their school placement in the first year of the new mechanism. Similar arguments for the flexibility provided by multiple offer systems are raised by critics of the National Residency Matching Program. See, e.g., Agarwal (2012).

Pan London scheme is based on the "equal preferences" algorithm, which is a version of the student-proposing deferred acceptance algorithm used in New York City (Pathak and Sönmez, 2011). A few years later, in 2012, officials in the Recovery School District in New Orleans also coordinated disparate processes for applying to a school within a single application form, citing the need to eliminate a "frustrating, ad-hoc enrollment process handled at individual campuses" (Vanacore, 2012).

We find compelling evidence that the new mechanism resulted in a 40% increase in enrollment in assigned school. The increase in take-up with the new mechanism is a widespread phenomenon, across students from each borough, racial group, and baseline achievement categories. The changes are greatest for white students, Staten Island residents, and those with high baseline achievement. In addition, 8.5% of New York City students left the district after submitting an application, while only 6.4% left under the new mechanism with the largest reaction coming from white and high-achieving students.

The most significant difference between assignments under the two mechanisms is how far students have to travel to their assigned school. In the old mechanism, the average travel distance was 3.45 miles, but this increases by 20% to 4.14 miles in the new mechanism. While most student groups travel further to attend their assigned school, they do so because they prefer those schools. Our preference estimates reveal that participants have substantially heterogenous tastes, trading off proximity and school quality. Students do not rank the closest school to their home, but prefer schools that are closer all else equal. Higher achieving students prefer higher achieving schools, though all students rank better performing schools higher on their rank order list. The estimates across our demand models are remarkably consistent across the many variations we consider, with almost no difference in the amount that families care for distance in our four main specifications.

The amount that students prefer their new school more than compensates for the greater distance they have to travel to attend the school, suggesting that the new mechanism has made it easier for students to express and obtain a choice they want. Ignoring the greater travel time, the improvement associated with the matched school is equivalent to about 0.96 miles; net of the greater travel distance, the average improvement is 0.25 miles. Students across boroughs, racial groups, and the spectrum of baseline achievement all obtain a more preferred assignment on average from the new mechanism. Given that some students eventually must attend less desirable schools under any mechanism, this pattern suggests the new mechanism was a net improvement for most students. The only student group who does worse on average are Special Education students, though the difference is only equivalent to an increase of 0.09 miles.

The reforms of the school choice mechanism in New York City also raise a number of design questions that we examine using our preference estimates. The new mechanism is 2.74 miles in equivalent utility below the utilitarian optimal assignment given only resource constraints on the number of school seats. Even though the new mechanism is based on the student-proposing deferred acceptance algorithm, it does not produce a student-optimal stable matching because some schools do not have strict orderings over students. In a student-optimal stable matching, student welfare improves by an equivalent of 0.10 miles. A Pareto efficient matching is equivalent to an improvement of 0.33 miles. Unfortunately, these allocations cannot be implemented without affecting the incentive properties of the mechanism and mechanisms that implement these allocations in equilibrium are not known (Abdulkadiroğlu et al., 2009; Erdil and Ergin, 2008; Kesten, 2010; Kesten and Kurino, 2012). Nonetheless, these counterfactuals show that the welfare gain from centralization is comparable to those from relaxing constraints imposed by the new mechanism's algorithm.

The literature on school choice is immense and not easily summarized. We share a focus with papers interested in understanding how choice impacts assignment and sorting of students (Epple and Romano, 1998; Urquiola, 2005), rather than the competitive effects of choice on student achievement (Hoxby, 2003; Rothstein, 2006). This study contributes to a growing literature on centralized admissions to college (Balinski and Sönmez, 1999) and K-12 public schools (Abdulkadiroğlu and Sönmez, 2003). A number of recent papers use micro data from assignment mechanisms to understand school demand (Hastings et al., 2009; He, 2012), but this information comes from variants of the Boston mechanism, a highly manipulable mechanism where parents have a strong incentive to rank schools strategically. An important precursor to our study is Niederle and Roth (2003b), but the rich data we have in this study allow us to examine the consequences of sorting, quantitatively evaluate student welfare, and assess alternative design choices. Finally, our work is complementary to theoretical papers that have investigated other assignment mechanisms using simulated data (Abdulkadiroğlu et al., 2012; Erdil and Ergin, 2008; Kesten, 2010), or have used submitted ordinal student preferences to examine theoretical points without relating them to empirically-grounded cardinalizations of utility (Abdulkadiroğlu et al., 2009; Pathak and Sönmez, 2008).

The paper proceeds as follows. Section 2 provides additional details on high school assignment in New York City. Section 3 describes our data. Section 4 reports on exit and enrollment and describes features of the two assignments. In Section 5 we outline our empirical strategy, while Section 6 reports estimates of the distribution of student preferences. Section 7 uses the demand estimates to compare student welfare across the decentralized and centralized assignments and to quantitatively assess tradeoffs in design choices. The last section concludes.

1.2 High School Choice in NYC

Forms of high school choice have existed in New York City for decades. Students have been able to apply for the city's venerable elite exam schools since the early part of the 20th century (Abdulkadiroğlu et al., 2011). In the period before 2002, schools in New York were organized into local community districts, and within-district choice options varied considerably among districts. An early well-known example was the District 4 choice plan in East Harlem, which attracted the attention of education reformers throughout the nation (Meier, 2002). In 1993, the city adopted open enrollment across all 32 community school districts (Schneider et al., 2002). The most significant change to high school choice took place upon Mayor Michael Bloomberg's election in 2001. Prior to his arrival, schools were overseen by an appointed Board of Education with decentralized control at the borough level. In June 2002, the state authorized mayoral control allowing for the establishment of the New York City Department of Education (NYC DOE), headed by Chancellor Joel Klein.

In the years before 2002, high school assignment in New York City featured a hodge-podge of choice options mostly controlled by borough-wide high school superintendents. While the choice system was administered centrally, significant admissions power resided with the schools because they could directly enroll students. Across the city, specific programs existed within schools, with curricula ranging from the arts to sciences to vocational training. It was not uncommon for multiple program types to exist within the same school, a feature that continues today. The set of school programs that developed at the time shaped the set of school options available during the time period of our study. The 2002-03 High School directory describes seven types of programs, categorized according to their admissions criteria.

At Specialized high schools, such as Stuyvesant and Bronx Science, there is only one type of program, which admits students by admissions test performance on the Specialized High Schools Admissions Test (SHSAT). Audition programs interview students for proficiency in specific performing or visual arts, music, or dance. Screened programs evaluate students individually using an assortment of criteria including a student's final 7th grade report card grades, reading and math standardized scores, attendance and punctuality, interview, essay or additional diagnostic tests. Educational Option programs also evaluate students individually, but for half their seats. The other half is allocated by lottery. Allocation of seats in each half targets a distribution of student ability: 16 percent of seats should be allocated to high performing readers, 68 percent to middle performers, and 16 percent to low performers. Unscreened programs admit students by random lottery, while Zoned programs give priority to students who apply and live in the geographic zoned area of the high school. Limited Unscreened programs allocate seats randomly, but give priority to students who attend an information session or visit the school's exhibit at a city-wide High school fair conducted each admissions season. We categorize programs into three main groups: Screened programs which also include Testing and Audition programs, Unscreened programs, which also include limited Unscreened and Zoned programs, and the rest are Educational Option programs.

Throughout the last decade, considerable high school reform efforts involved closing and opening of high schools throughout the city. One important change involved an expansion of the number of small high schools, usually with fewer than 500 students. A big push for these small high schools came as part of the Children's First Initiative launched by Mayor Bloomberg and Chancellor Klein, supported by private institutions like the Gates Foundation. Most of these new schools replaced large high schools, though they were often housed at the same location. They are also commonly Limited Unscreened. In 2003, 32 additional new small schools opened, but in the first year of the centralized mechanism there were 69 new schools, 61 of which were classified as small schools. The following year had an even greater number of new schools created (Abulkadiroglu et al., 2012). Most of these schools are small and have less than 100 students per entering class. As a result, they have a relatively small impact on overall enrollment patterns. Many of these new schools were announced only after the high school directory was printed, so we will have to pay special attention to our preference estimates for these schools.

1.2.1 Decentralized Admissions in 2002-2003

Admissions to the Specialized High Schools and the LaGuardia High School of Music & Art and Performing Arts have been traditionally administered as a separate process from the assignment to regular high schools, and this did not change with the new mechanism.[3] In 2002-03, nearly 30,000 rising high school students applied for about 5,000 spots at one of the five specialized high schools or LaGuardia by taking the SHSAT in late October. On the day of the exam, students rank these schools in order of preference.

About 80,000 students who are interested in regular high schools visit schools and attend city-wide high school fairs before submitting their preference in early December. In 2002-03, students could apply to at most five regular programs in addition to the Specialized high schools. Programs receiving a student's application were able to see the applicant's entire preference list, including where their program was ranked. Programs then decided whom to accept, place on a waiting list, or reject. Applicants were sent a decision letter from each program they had applied to and some obtained more than one offer. Students were allowed to accept at most one admission and one wait-list offer. After receiving responses to the first letters, programs with vacant seats could make new offers to students from waiting lists. After replies to the second letter were received, a third letter with the new offers was sent. New offers did not necessarily go to wait-listed students in a predetermined order. Remaining unassigned students were assigned their zoned programs or

[3]The 1972 Hecht-Calandra Act is a New York State law that governs admissions to the original four Specialized high schools: Stuyvesant, Bronx High School of Science, Brooklyn Technical, and Fiorello H. LaGuardia High School of Music and Performance Arts. City officials indicated that this law prevents including these schools within the common application system without an act of the state legislature.

assigned via an administrative process, where the central office tried to place kids as close to home as possible. Finally, there was an informal appeals process.

Three features of this assignment scheme motivated the NYC DOE to abandon this process in favor of a new mechanism. First, there was inadequate time for offers, wait list decisions, and acceptances to clear the market for school seats. This feature of congestion is seen in other decentralized matching processes such as in the market for clinical psychologists (Roth and Xing, 1997). DOE officials reported that in many cases high achieving students received acceptances from all of the schools they applied to, while many received none (Herszenhorn, 2004). Comments by the Deputy Schools Chancellor summarize the frustration expressed with the system: "Parents are told a school is full, then in two months, miracles of miracles, seats open up, but other kids get them. Something is wrong" (Gendar, 2000).

Second, some schools awarded priority in admissions to students who ranked them first in their application. The high school directory advises that when ranking schools, students should "... determine what your competition is for a seat in this program." This recommendation complicates the ranking decision for parents. Listing such a school first would improve the likelihood of an offer from that school at the risk of rejection by one of their lower choice schools that took students' rankings into account.

Third, a number of schools managed to conceal capacity to fill seats later on with better students. For example, the deputy chancellor of schools defended the new plan stating, "before you might have a situation where a school was going to take 100 new children for ninth grade, they might have declared only 40 seats, and then placed the other 60 outside the process" (Herszenhorn, 2004). Overall, critics alleged that the old mechanism disadvantaged low-achieving students, and those without sophisticated parents (Hemphill and Nauer, 2009).

1.2.2 Centralized Admissions after 2003

The new mechanism was designed with input from economists (Abdulkadiroğlu et al., 2005). When publicizing the new mechanism, the DOE explained that its goals were to utilize school places more efficiently and to reduce the gaming involved in obtaining school seats (Kerr, 2003). The first round involves students applying to Specialized high schools when they take the SHSAT, submitting a ranking of Specialized schools as in previous years. Offers are produced according to a serial dictatorship with priority given by students' admissions test scores.

In the main round, students interested in regular schools can rank up to twelve programs in their application, which are due in November. The DOE advises parents: "You must now rank your 12 choices according to your true preferences" because this round is built on Gale and Shapley (1962)'s student-proposing deferred acceptance algorithm. Schools with programs that prioritize applicants based on auditions, test scores or other criteria are sent lists of students who ranked the school, but these lists do not reveal where in the preference lists they were ranked. Schools return orderings of applicants to the DOE Enrollment office. Schools which prioritize applicants using geographic or other criteria have those criteria applied by the central office. That office uses a single lottery to break ties amongst students with the same priority, generating a strict ordering of students at each school.

Assignment is determined by the student-proposing deferred acceptance algorithm with student preferences over the schools, school capacities, and school's strict ordering of students as parameters. The algorithm works as follows:

> Round 1: Each student applies to her first choice school. Each school rejects the lowest-ranking students in excess of its capacity, with the rest provisionally admitted (students not rejected at this step may be rejected in later steps.)
>
> Round $\ell > 1$: Students rejected in Round ℓ-1 apply to their next most preferred school (if any). Each school considers these students *and* provisionally admitted students from the previous round together, rejecting the lowest-ranking students in excess of capacity, producing a new provisional admit list (again, students not rejected at this step may be rejected in later steps.)

The algorithm terminates when either every student is matched to a school or every unmatched student has been rejected by every school she has ranked. A student obtains at most one offer or is unassigned. The algorithm is run with all students in February. In this first round, only students who receive a Specialized high school offer receive a letter indicating their regular school assignment, and are asked to choose one. After they respond, students that accept an offer are removed, school capacities are adjusted and the algorithm is run again with all remaining students. All students receive a letter informing them of their assignment or if they are unassigned after the main round.

Unassigned students from the main round are provided a list of programs with vacancies, and are asked to rank up to twelve of these programs. In 2003-04, the admissions criteria at the remaining school seats are ignored in this supplementary round. Students are ordered by a single random list, and the student-proposing deferred acceptance algorithm is run with this ordering in place at each school. Students who remain unassigned in the supplementary round are assigned administratively. In the first year of the new centralized mechanism, the DOE also conducted an informal appeals process where students could request a new placement if they were dissatisfied with their assignment. These appeals were manually processed on a case-by-case basis. In the first year of implementation, the new mechanism resulted in about 8,000 unassigned students after the main round, and 3,000 administratively assigned students.

1.3 Data and Market Description

The NYC DOE provided us with several data sets for this study each linked by a unique student ID number: information on student choices and assignments, student demographics, and October student enrollment. The choice file is maintained by the Enrollment office (formerly known as the Office of Student Enrollment and Planning Operations or OSEPO), which runs high school admissions. For 2002-03, the OSEPO files only record a student's program assignment at the conclusion of the assignment process, and not their initial rank order list or sequences of offers and rejections from the decentralized mechanism. As a result, we cannot re-create the decentralized assignment, and only know each applicant's final assignment. For 2003-04, the OSEPO files contains students' choice schools in order of preference, priority information for each school, and assignments at the end of each of the rounds. The student demographic file contains information on sending school, home address, gender, race, limited English proficiency, special education status, and performance on middle school standardized tests. We use MS MapPoint to compute the travel distance between each student and school, and ArcGIS with StreetMap USA to geocode each student to a census tract corresponding to his or her address. In New York, high school students who live within 0.5 miles of a school are not eligible for transportation. If a student lives between 0.5 and 1.5 miles the Metropolitan Transit Authority provides them with a half-fare student Metrocard that works only for bus transportation. If they reside 1.5 miles or greater, they obtain full fare transportation with a student Metrocard that works for subways and buses and is issued by the school transportation office.

Our analysis sample makes two restrictions. First, since we do not have demographic information for private school applicants, we restrict the analysis to students in NYC's public middle schools at baseline. Second, we focus on students who are not assigned to Specialized high schools because that part of the assignment process did not change with the new mechanism. Most students prefer a Specialized high school to a mainstream high school, limiting the potential for interaction with the main round to differentially influence assignments as the mechanism changed. Given these restrictions, we have two main analysis files: the *welfare* sample and the *demand* sample.

The welfare sample is used for comparisons of the assignment across the two mechanisms. We construct the welfare sample as the largest set of students to be assigned through the high school assignment mechanism to a school that exists as of the time of the printing of the high school directory. Any non-private applicant in the OSEPO files who does not opt for their current school and is assigned is a member of the welfare sample in 2002-03 and 2003-04. Columns (1) and (2) of Table 1.1 summarize student characteristics in our welfare samples across years. 2,500 fewer students are involved in the 2003-04 sample, which is mainly due to the students assigned to schools created after the printing of the high school directory.

New York is the nation's largest school district, and like many urban districts it is majority low-income and non-white. More than a third of students black and a third are Hispanic, while about 10% of students are Asian. The greatest number of public school students are from Brooklyn, followed by Queens and the Bronx, which each account for roughly one quarter of the students. Manhattan and Staten Island, areas with high private school

Table 1.1: Descriptive Statistics for Students

Student sample for:	Mechanism Comparison		Demand Analysis
	Decentralized Mechanism (1)	Centralized Mechanism (2)	Centralized Mechanism (3)
Total Number of Students	70358	66921	69907
Manhattan	12.5%	11.8%	12.0%
Brooklyn	31.9%	34.1%	33.3%
Queens	25.0%	24.8%	24.7%
Bronx	23.7%	23.3%	23.7%
Staten Island	6.9%	6.0%	6.3%
Asian	10.6%	10.9%	10.6%
Black	35.4%	35.7%	35.7%
Hispanic	38.9%	40.4%	40.3%
White	14.7%	12.6%	13.0%
Other	0.4%	0.4%	0.4%
Female	49.4%	49.0%	49.0%
Limited English Proficiency	13.1%	12.6%	12.6%
Special Ed	8.2%	7.9%	7.5%
Taken the SHSAT	22.4%	24.3%	23.9%
Mean Household Income	44184	43783	43823
Std. Household Income	19642	19539	19542
Mean Family Income	48345	47799	47860
Std. Family Income	25592	25272	25289

Notes: Summary of characteristics of students samples. Decentralized mechanism refers to 2002-2003 mechanism; centralized mechanism refers to the deferred acceptance mechanism adopted in 2003-2004. Income characteristics are from the student's census-block means for the 2000 census.

penetration, account for a considerably smaller share of students at 12 and six percent, respectively. Student characteristics of participants are similar across years, suggesting the attributes of participants did not change substantially as a result of the new mechanism.

The demand sample contains participants in the main round of the new mechanism in the OSEPO files. The school choices expressed by this group of students represent the overwhelming majority of students. Since we do not observe student choices in the decentralized mechanism, it does not have a corresponding demand sample. Among the set of main round participants, we exclude a small fraction of students who are classified as the top 2 percent because these students are guaranteed a school only if they rank it first and this may distort their incentives to rank schools truthfully. Additional details on the sample restrictions are in the data appendix. Table 1.1 also shows that the student characteristics are similar across the sample used for comparing mechanisms and demand samples. The comparability of the two samples is an important condition for us to use the demand sample to make statements about the welfare samples for both years.

Data on schools were taken from the 2003-04 report card files provided by NYC DOE. Information on programs come from the official NYC High School Directories made available to students before the application process. Table 1.2 summarizes school and program characteristics across years. There is an increase in the number of schools from 215 to 234, and a corresponding decrease in the average number of students enrolled per school of about 40 students. This fact is driven by the replacement of some large schools with smaller schools that took place concurrently in 2003-04 described above. Despite this increase, there is little change in the average achievement levels of schools and their demographic composition. Moreover, slightly more than half of the teachers have taught for less than two years, a pattern which holds for both years.

Students in New York can choose among roughly 600 programs throughout the city. The menu of program choices is immense, and programs vary substantially in focus, post-graduate orientation, and educational philosophy. For instance, the Heritage School in Manhattan is an Educational Option school where the arts play a substantial role in learning, while Townsend Harris High School in Queens is a Screened program with a rigorous humanities program making it among the most competitive in the city. Using information from high school directories, we identify each program's type, language orientation,

Table 1.2: Descriptive Statistics for Schools and Programs

	Decentralized Mechanism (1)	Centralized Mechanism (2)
A. Schools		
N	215	235
Students assigned per school	327.2	284.8
High English Achievement	19.1	19.3
High Math Achievement	10.2	10.0
Percent Attending 4yr college	47.8	47.2
Percent Inexperienced Teachers	54.7	55.6
Percent Free or Reduced Price Lunch	62.5	62.6
Attendance Rate (on 100)	85.5	85.7
Asian	8.7	8.6
Black	38.5	38.4
Hispanic	41.9	42.1
White	10.9	10.9
B. Programs		
N	612	558
Screened	233	208
Unscreened	63	119
Education Option	316	119
Spanish Language	27	24
Asian Language	10	9
Other Language	6	7
Arts	80	80
Humanities	89	93
Math and Science	53	60
Vocational	55	59
Other Specialties	163	162

Notes: Data availability on school characteristics varies, as described in data appendix. High Math and High English achievement are the fraction of student that scored more than 85% on he Math A and English regents tests in 2003-2004, respectively. Teachers that have taught for less than two years are considered inexperienced.

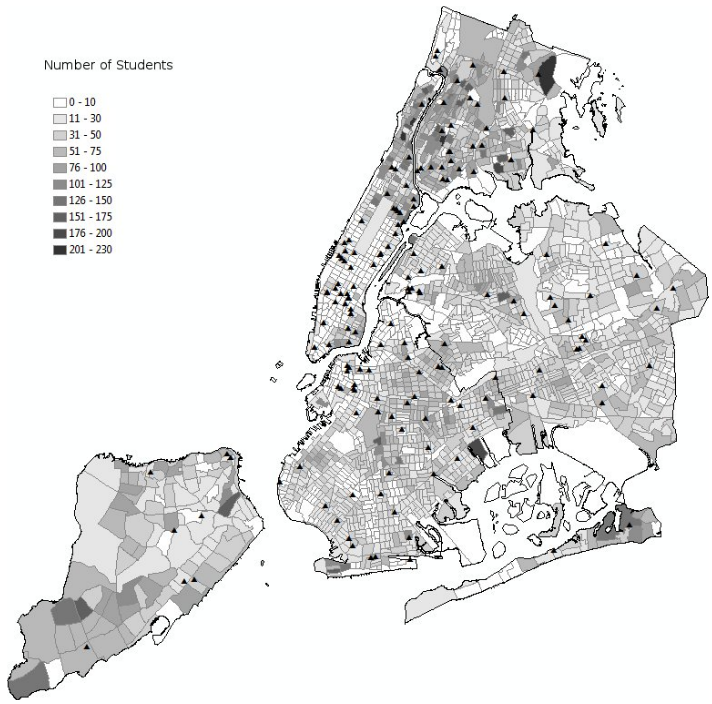
and speciality. With the new mechanism, there are more Unscreened programs and fewer Educational Option programs, a change partly due to the conversion of some Educational Option programs to Unscreened programs, since half of the seats at Educational Option programs are allocated via random lottery as in an Unscreened program. Another reason for this increase is that the new small schools are often categorized as limited Unscreened. We code language-focused programs as Spanish, Asian, or Other, and categorize program specialities into Arts, Humanities, Math and Science, Vocational, or Other. Not all programs have specialties, though about 70% fall into one of these classes. (Details on our classification scheme are in the data appendix). The menu of language program offerings or program specialities changes little across years.

Given the abruptness of the announcement of the new mechanism, it seems reasonable that there was little scope for participants to react to it by either opting out or in to the mechanism. Likewise, it does not appear that there is a large change in the residential locations of students. The distribution of students throughout the city looks nearly identical to that in the new mechanism, shown in the map in Figure 1.1. Moreover, though there are relatively small changes in the set of school options, these mostly involve an increase in the number of Unscreened programs, and a decrease in the number of Educational Option programs, and these two program types are similar. These facts together motivate our approach of using preference estimates from choices expressed in 2003-04 to measure student welfare from the final assignments in 2002-03, and to attribute those to changes in the assignment mechanism rather than changes in the attributes of student participants or the menu of school options.

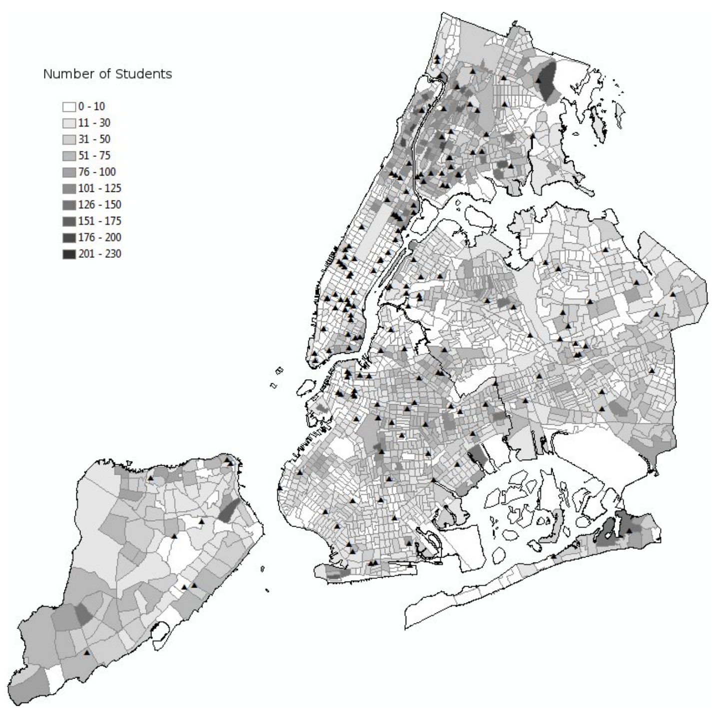
1.4 Descriptive Evidence on Mechanism Performance

1.4.1 Exit and Noncompliance

We begin by analyzing the decision of a student not to enroll in a public school though she is assigned a seat in the NYC public school system. For instance, after receiving an assignment, a student may opt for a private school, leave New York, or even drop out. A student is coded as an "exit" if he or she participates in the choice process, but does not



(a) 2002-2003



(b) 2003-2004

Figure 1.1: School Locations and Students by New York City Census Tract

enroll in any New York City public school as of October of the next school year. If students are dissatisfied with their assignments, we predict they are more likely to exit. Columns (1) and (2) in Table 1.3 report the percentage of students who are assigned a NYC public school, but matriculate at a school outside the NYC DOE system under the decentralized and centralized system, respectively. New York, like many urban districts, has substantial student turnover: 8.4% of students participate in the choice process, but do not enroll in a New York City public school in 2002-03. However, about 1,600 fewer students (or 2%) exit under the centralized system.

Changes in exit rates vary considerably by student demographics, with the largest changes for white students (45%) and students from Staten Island (63%). It is worth noting that these are relatively small subgroups: only about 14% of students in New York City are white, and about 6% are from Staten Island. However, exiting may be a more feasible option for these groups. The importance of the ability to exit is also seen when we compare students across achievement and income lines. Exit rates decrease by 40% for high baseline achievement students, but only 6% for low baseline students. Likewise, students who reside in the highest income quartile census block group are 38% less likely to exit, while those in the lowest income quartile are 10% less likely to exit.

The evidence that exit rates drop the most for students who likely have good outside options weighs against the view that multiple-offers systems are important for keeping families with good outside options in the public district. Rather, these groups are less likely to exercise their outside option in the coordinated single-offer system. Moreover, the reduction in exit rates for these groups does not come at the expense of other student groups, as drop in exit rates is observed at differing levels across all races, residential locations, baseline achievement levels and income groups. This fact provides the first indication that changes in allocation systems are not necessarily zero-sum.

It is worth emphasizing however that changes in exit rates can happen for a number of reasons including different economic conditions or demographic trends in the city, and therefore these differences cannot be solely attributed to the new mechanism. Perhaps a better measure of participant satisfaction is whether a student enrolls at her assigned school, or she has switched by October of the next school year. Families may switch schools after their final assignments are announced but before the school year starts for a number of

Table 1.3: Exit and Enrollment across Mechanisms

	Exit from NYC Public Schools		Enrolled in Unassigned School	
	Decentralized Mechanism (1)	Centralized Mechanism (2)	Decentralized Mechanism (3)	Centralized Mechanism (4)
All Students	8.5%	6.4%	18.4%	11.4%
Female	8.2%	6.2%	18.1%	11.3%
Asian	6.0%	4.0%	17.5%	6.6%
Black	7.3%	6.4%	19.0%	12.5%
Hispanic	6.1%	5.5%	18.2%	12.6%
White	19.2%	10.4%	18.1%	8.3%
Manhattan	7.2%	6.2%	17.3%	10.5%
Brooklyn	6.4%	5.1%	17.2%	11.2%
Queens	8.5%	6.6%	21.1%	10.7%
Bronx	9.2%	8.0%	19.1%	14.3%
Staten Island	18.5%	6.8%	14.2%	6.1%
High Achievement	12.3%	7.2%	15.1%	7.0%
Mid Achievement	7.5%	5.9%	16.3%	10.4%
Low Achievement	4.9%	4.7%	18.6%	15.1%
Lowest Income Quart.	5.8%	5.2%	18.6%	13.6%
Second Income Quart.	6.8%	6.0%	18.7%	11.8%
Third Income Quart.	8.2%	6.5%	17.8%	11.3%
Highest Income Quart.	13.1%	7.7%	18.7%	8.7%

Notes: Means. Exit is enrollment in a school outside the NYC Public School System conditional on assignment. Enrollment in unassigned school is conditional on enrollment in a NYC Public School. Students with a score under the 25th percentile (over the 75th percentile) in the middle school math examination are categorized as having a Low Achievement (High Achievement). Decentralized mechanism refers to 2002-2003 mechanism; centralized mechanism refers to the deferred acceptance mechanism adopted in 2003-2004. Income characteristics are from the student's census-block means for the 2000 census.

reasons, including moving to different parts of the city. In the decentralized assignment system, principals had greater discretion to enroll students and the DOE officials quoted above alleged that students with sophisticated parents might just show up at a school in the fall and ask for a place. This strategy of getting into a school became more difficult with centralized control of assignment. In the decentralized mechanism, for students who did not exit, 18% of students, or 11,700 students, enrolled in a school other than their assigned school. However, in the centralized mechanism 4,600 more students enroll in their assigned school, a decrease in the non-compliance rate of 40% with the centralized mechanism. Whites and Staten Island residents are more likely to take up their assigned school, but unlike the pattern for exits, they are more comparable to Asians and students living in Queens. As with exits, however, high baseline students and high income students are more likely to enroll in their assigned school under the new mechanism.

1.4.2 Sorting and Stratification

Much literature on school choice examines whether the presence of choice options impacts the sorting of students into schools across the district. For instance, Epple and Romano (1998)'s equilibrium model of school vouchers has a hierarchy of schools with student bodies entirely stratified by income and ability. In the last section, we saw that high baseline and income students were less likely to exit, so it is possible that they remained in New York City's public schools because they are more likely to attend school with peers with similar attributes. If adopting a centralized choice mechanism leads to an increase in student's ability to express choice, it is possible that schools have become more stratified with high achieving students more likely to attend school with one another or certain demographic groups clustered together. In the context of inter-district choice, Clotfelter (1999) argues that school-level peer groups will be impacted by the presence of choice, while Hoxby (2003) finds that racial heterogeneity is not impacted by the number of districts. Urquiola (2005) reports that inter-district negatively impacts the racial heterogeneity and reduces private school enrollment. In this section, we examine whether the centralization of choice impacts stratification.

Under both mechanisms, we compute the racial composition and baseline achievement

Table 1.4: Comparing Assignments

	All Students		Racial Composition							
			Asian		Black		Hispanic		White	
	Old (1)	New (2)	Old (3)	New (4)	Old (5)	New (6)	Old (7)	New (8)	Old (9)	New (10)
Panel A: Peer Characteristics										
Asian	10.6%	10.9%	21.7%	22.7%	7.4%	7.8%	9.1%	9.0%	14.3%	15.7%
Black	35.4%	35.7%	24.7%	25.4%	49.5%	48.8%	30.9%	30.9%	21.4%	22.6%
Hispanic	38.9%	40.4%	33.5%	33.2%	33.9%	35.0%	49.8%	50.8%	26.0%	28.5%
White	14.7%	12.6%	19.8%	18.2%	8.9%	8.0%	9.8%	8.9%	37.9%	32.8%
High	24.0%	23.4%	33.1%	34.4%	20.8%	20.3%	20.1%	19.8%	35.6%	34.6%
Mid	40.0%	47.7%	36.1%	42.5%	41.3%	49.3%	41.0%	48.9%	36.9%	43.9%
Low	22.5%	21.0%	16.4%	15.5%	25.1%	22.8%	24.6%	22.8%	15.2%	15.1%
No Score	13.5%	7.8%	14.4%	7.5%	12.8%	7.6%	14.3%	8.5%	12.3%	6.3%
Manhattan	12.5%	11.8%	9.9%	8.3%	10.6%	9.8%	17.3%	16.1%	6.1%	6.8%
Brooklyn	31.9%	34.1%	29.9%	32.7%	43.4%	45.6%	21.0%	23.3%	34.9%	37.4%
Queens	25.0%	24.8%	49.1%	47.3%	19.7%	20.2%	23.5%	23.4%	24.4%	23.3%
Bronx	23.7%	23.3%	6.6%	7.2%	23.2%	21.5%	35.1%	34.0%	6.9%	7.3%
Staten Island	6.9%	6.0%	4.5%	4.4%	3.2%	2.9%	3.1%	3.2%	27.8%	25.3%
Panel B: Characteristics of Assigned School										
Mean Distance Traveled (mi)	3.36	4.05	2.67	3.30	3.73	4.47	3.51	4.18	2.54	3.10
Median Distance Traveled (mi)	2.25	3.04	1.71	2.19	2.68	3.55	2.37	3.24	1.79	2.13
High English achievement	19.2	19.8	26.0	27.9	16.8	17.3	15.8	16.4	28.4	29.5
High Math achievement	11.2	11.4	18.3	19.6	8.8	9.1	9.2	9.4	16.8	17.0
Percent attending 4yr college	47.4	48.8	53.6	54.9	44.4	46.0	44.4	46.4	57.4	58.1
Free/Reduced Lunch	55.6	55.3	44.1	43.0	57.3	56.1	64.2	63.6	37.0	37.8
Panel C: Ranking and Assignment										
Number of choices		7.76		6.60		8.67		8.30		4.55
Choice received		3.00		2.97		3.20		3.07		2.31

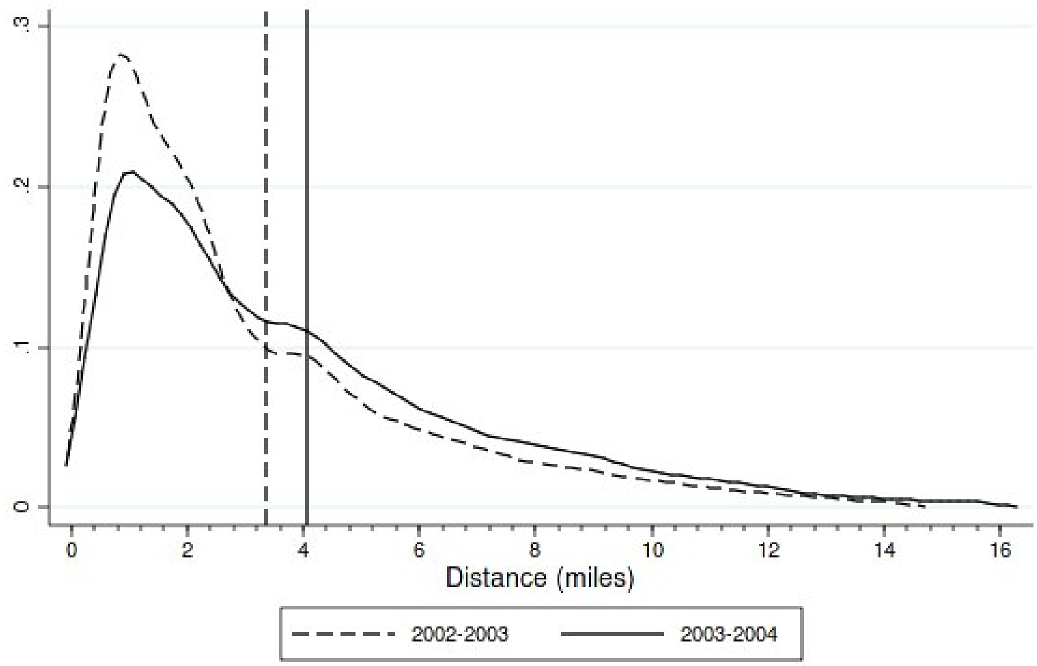
Notes: Columns denote characteristics of students and rows denote the average characteristics of peers assigned to the same school. Columns (5) and (10) use all students. Students with a score under the 25th percentile (over the 75th percentile) in the middle school math examination are categorized as having a Low Baseline Score (High Baseline Score).

of a typical applicant's peers in her assigned school. Across these measures, there is little evidence that the racial composition of peers changes significantly across the two mechanisms in Table 1.4. For instance, 36% of applicants are black in both the decentralized and centralized mechanism. In the decentralized mechanism, 50% of the peers of a typical black applicant are themselves black, while under the centralized mechanism only 49% are. Asians and Hispanics are assigned schools with 1% more Asians and 1% more Hispanics, respectively. The most noticeable change appears for white students, who are assigned to schools with a smaller percentage of whites (38% vs 33%).

Turning to baseline achievement, the sorting patterns of students also does not change considerably with the new mechanism. Though there is a difference for those with middle Math baseline scores, this is driven by greater availability of middle school scores. Middle-tier math students are assigned schools with a higher percentage of students with middle baseline scores by the centralized mechanism (42% vs 50%), but there is an increase in the percentage of middle score students in the population (40% vs 48%) from old and new mechanisms in columns (1) and (2), which is mainly due to the decrease in the percentage of students with no scores (14% vs 8%). Indeed, the percentage of middle scorers among students' peers increase at all achievement levels. Therefore, columns (11)-(16) do not offer evidence for changes in stratification across achievement levels.

We next examine sorting patterns via the characteristics of the assigned school in Panel B. Overall, achievement levels or the fraction of students attending a four year college do not change significantly between the two mechanisms. Any difference in the assigned school's attributes by race seem relatively minor compared with year-by-year fluctuations in these school-level measures. Indeed, 17.6% of high baseline math students attend a school with high Regents Math achievement (measured by the fraction scoring 85% or higher) under the centralized mechanism, while 16.7% do so in the decentralized mechanism. There are also virtually no differences in the poverty status of assigned schools for racial and baseline achievement subgroups. It is possible that sorting patterns have changed in ways not captured by the school characteristics in Table 1.4, so when we estimate preferences for schools, it will be important to allow for school-specific unobservable attributes.

The most pronounced difference between assignments involves the amount of distance students have to travel to attend school. In Figure 1.2, we report the overall distribution of



Mean (median) travel distance is 3.36 (2.25) miles in 2002-2003 and 4.05 (3.04) miles in 2003-2004. Top and bottom 1% are not shown in figure. Line fit from Gaussian kernel with optimal bandwidth

Figure 1.2: Distribution of Distance to Assigned School in 2002-2003 and 2003-2004

distance travelled by students in both mechanisms. New York City spans a large geographic range, with nearly 45 miles separating the southern tip of Staten Island with the northern most points of the Bronx, and 25 miles traveling from the western edge of Manhattan near Washington Heights to Far Rockaway at the easternmost tip of Brooklyn. The closest school for a typical student is 0.86 miles from their home, but students in 2002-03 on average travelled 3.45 miles from their home address to their assigned school. In the new mechanism, the average distance is 4.14 miles. The medians also increase from 2.45 to 3.21 miles.

Students across racial, baseline achievement and income levels are travel further to their assigned school. Table 1.4 shows a increase in distance for Asians, blacks, whites, Hispanics and whites, with the smallest increase for white students (0.54 miles), and the largest increase for blacks (0.76 miles). The increase in travel distance to assigned school is also uniform across high, middle and low baseline achievement levels. The one exception to this pattern involves comparisons across boroughs. Although not reported in this table, students from Manhattan on average experience no increase in the their travel distance to school, while students from the other boroughs do. The largest impact is for students in the Bronx and Queens (0.97 and 1.09 miles on average). The increase in distance due to a centralized mechanism is consistent with the pattern documented by Niederle and Roth (2003b) in their study of the gastroentrology labor market. Mobility in their study is defined as the percentage of gastroenterology fellows who change hospitals (or cities or states) after finishing their previous training (usually in internal medicine) and starting the gastroenterology fellowship. They report mobility increases at the hospital, city, and state level with the introduction of a centralized match.

Based on their submitted preferences, all else equal, students strongly prefer attending a school closer to home. The fact that students in the new mechanism are assigned to schools further from home might suggest that the new mechanism led to assignments that are worse on average than the old mechanism. On the other hand, students may prefer schools outside of their neighborhood because they are a better fit for the student. Given this fact, we must weigh the greater travel distance in the new mechanism against the attributes of the school. Students do not rank the closest school to their home, however, and instead trade off school attributes with proximity. Our next task is to quantify how students evaluate distance relative to school attributes such as size, demographic composition, and average test scores in their submitted preferences.

1.5 Measuring Student Preferences

1.5.1 Student's School Choices and Assignments

Before providing details on student's choices, we first examine some descriptive patterns of student assignments. In Panel C of Table 1.4, we report the average number of choices and choices received from the demand sample. On average, applicants rank eight school programs and receive their third choice.[4] Asian and white students rank fewer schools (7.3 and 6.6 schools, respectively) than black and Hispanic applicants (8.1 and 8.2 schools, respectively). Despite ranking fewer schools, Asians and whites on average receive an assignment slightly better than their third choice. High baseline students rank fewer schools than middle and low baseline students, and both obtain their third choice on average.

What makes a families rank particular schools and where do they obtain information about school options? In surveys, parents state that academic achievement, school and teacher quality are the most important school characteristics (Schneider et al., 2002). In New York, families obtain information about high schools and programs from many sources. Guidance counselors, teachers, peers, and other families in the neighborhood all provide input and recommendations. The official repository of information on high schools is the NYC High School directory which includes information about school size, advanced course offerings, Regents and graduation performance, as well as the school's address, and closest bus and subway. The directory also includes a paragraph description of each program together with a list of extracurricular activities and sports teams. Families learn about schools at High school fairs that are held on Saturdays and Sundays in the fall in each borough, from individual school open houses, from online school guides such as insid-

[4]Table A.5 provides additional information on school assignments. 34.0% of students receive their top choice, 15.5% receive their second choice, and 2.4% receive a choice ranked 10th, 11th, or 12th. 14.4% of students are asked to participate in the supplementary round because they are unassigned in the main round.

eschools.org, and from books about high schools such as Hemphill (2007). Finally, local newspapers such as the New York Post and New York Daily News regularly publish rankings of high schools, using publicly available information from the New York State Report Cards.

Even though the school ranking decision involves evaluating many different alternatives, the aggregate distribution of preferences displays some consistent regularities: students prefer closer and higher quality schools, shown in Table 1.5. The first row of the table shows that only 23% of applicants rank 12 school choices; the majority 9 or fewer choices and nearly 90% rank at least three choices. A typical student's top choice is 3.6 miles away from her home. Given that the closest school is 0.86 miles away from the average student, this means that students do not rank their closest school first, but rather consider schools further away from home. An average student's first choice is nearly 0.40 miles closer than her second choice, and their second choice is about 0.25 miles closer than her third choice. Despite the possibility that other school characteristics influencing preferences, distance to a ranked school increases monotonically until the 9th choice, which is 4.94 miles away on average. The distance to the 10-12th choices is lower than the 9th choice, but this may be due to changes in the composition of students who rank longer lists.

Lower ranked schools are not only farther away, but they also less desirable on other measures of school quality. We measure quality following the Regents performance information in the 2003-04 High School directory. The fraction of students scoring an 85 or higher on the English or Math Regents exam decreases going down rank order lists. The percent of students attending a 4-year college also decreases with rank, and the fraction of teachers classified as inexperienced increases. Each of these school quality measures are also monotonically related to rank. Finally, schools enrolling a higher share of white and Asian students tend to be ranked higher than schools with smaller shares of these student groups.

Using requests for individual teachers, Jacob and Lefgren (2007) find that parents in low-income and minority schools value a teacher's ability to raise student achievement more than in high-income and non-minority schools. This difference across groups motivates our investigation of ranking behavior by baseline ability and income quartile. Students who are low achievers care for distance nearly in the exact same way as students who

Table 1.5: School Characteristics by Student Choice in Centralized Mechanism

Choice	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th	12th
						A. Full Sample						
N	69907	94%	89%	83%	77%	69%	63%	56%	50%	43%	35%	20%
Mean Distance (miles)	4.43	4.81	5.05	5.21	5.38	5.49	5.59	5.63	5.65	5.58	5.43	5.12
Median Distance (miles)	3.51	3.95	4.20	4.37	4.57	4.63	4.72	4.77	4.78	4.71	4.59	4.24
High English Achievement	27.8	25.6	24.3	23.2	22.2	21.4	20.8	20.5	19.9	19.4	19.1	18.6
High Math Achievement	16.7	15.3	14.7	13.9	13.4	12.8	12.4	11.9	11.5	11.1	10.8	10.4
Percent attending 4yr college	53.7	54.1	52.8	51.7	50.9	50.1	49.3	49.3	49.0	48.3	47.6	47.5
Fraction Inexperienced Teachers	43.0	43.8	45.0	46.2	47.1	48.1	48.7	49.5	49.5	50.2	49.9	49.9
Free or Reduced Price Lunch	51.4	53.4	54.5	56.2	57.4	58.7	59.8	60.6	61.3	61.9	62.7	63.1
Attendance Rate	87.2	86.7	86.4	86.1	85.9	85.7	85.5	85.5	85.2	85.0	84.8	84.5
Asian	13.8	13.7	13.3	12.7	12.2	11.5	11.0	10.6	10.1	9.7	9.5	9.4
Black	32.6	34.3	35.1	36.0	36.7	37.5	38.2	38.4	38.7	38.9	38.9	38.1
Hispanic	34.4	35.2	35.9	37.0	37.9	38.9	39.6	40.2	40.8	41.5	42.2	43.5
White	19.1	16.7	15.7	14.4	13.3	12.2	11.2	10.8	10.4	9.8	9.4	9.0

Table 1.5: School Characteristics by Student Choice in Centralized Mechanism (Cont'd)

Choice	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th	12th
					B. Students with Low Baseline Scores							
Mean Distance (miles)	4.09	4.44	4.66	4.86	5.02	5.11	5.18	5.19	5.21	5.11	5.01	4.83
Median Distance (miles)	3.11	3.60	3.88	4.07	4.21	4.27	4.37	4.32	4.37	4.24	4.18	4.08
High English Achievement	20.1	19.4	18.7	18.4	18.1	17.9	17.6	17.7	17.4	17.1	16.6	16.6
High Math Achievement	10.9	10.9	10.5	10.1	10.0	9.7	9.6	9.4	9.4	9.1	8.8	8.8
					C. Students with High Baseline Scores							
Mean Distance (miles)	4.59	5.03	5.24	5.43	5.67	5.85	5.92	6.00	6.08	6.16	5.83	5.29
Median Distance (miles)	3.77	4.14	4.39	4.58	4.84	5.03	5.04	5.13	5.21	5.34	4.96	4.25
High English Achievement	39.4	34.0	32.0	29.8	28.0	26.4	25.2	24.6	23.7	22.8	22.5	21.5
High Math Achievement	26.0	21.4	20.5	19.1	18.2	17.3	16.4	15.6	15.2	14.2	13.9	12.8
					D. Lowest Income Quartile Students							
Mean Distance (miles)	4.35	4.58	4.74	4.89	5.02	5.06	5.11	5.17	5.17	5.08	4.93	4.73
Median Distance (miles)	3.51	3.86	3.98	4.13	4.27	4.29	4.37	4.35	4.39	4.32	4.21	4.06
High English Achievement	20.7	20.0	19.4	19.0	18.6	18.1	18.0	18.1	17.6	17.1	17.2	16.6
High Math Achievement	11.4	10.9	10.7	10.4	10.2	10.1	10.0	9.8	9.6	9.3	9.2	8.8
					E. Highest Income Quartile Students							
Mean Distance (miles)	4.60	5.14	5.45	5.67	5.87	6.04	6.20	6.31	6.41	6.43	6.30	5.81
Median Distance (miles)	3.56	4.12	4.44	4.67	4.90	5.02	5.07	5.28	5.38	5.36	5.21	4.57
High English Achievement	35.4	31.8	29.7	28.2	26.5	25.8	24.4	24.0	23.1	22.3	21.8	21.3
High Math Achievement	23.3	20.7	19.5	18.8	17.7	16.9	15.9	15.2	14.9	14.1	13.4	13.1

Notes: Data availability on school characteristics varies, as described in data appendix. High Math and High English achievement are the fractions of students scoring over 85% on the Math A and English regents in 2003-2004, respectively. Inexeperienced teachers are those who have been teaching less than 2 years. Students with a score under the 25th percentile (over the 75th percentile) in the middle school match examination are categorized as having a Low Baseline Score (High Baseline Score).

have high baseline achievement for their top choices. The average distance to the average first choice is 4.5 miles for both groups, while the 9th choice is 5.6 for low achievers and 5.9 miles for high achievers. High achieving students also tend to rank schools with high English and Math achievement, relative to low achievers, though both groups place less emphasis on achievement with further down their preference list. Similarly, students from lower income neighborhoods tend to put less weight on Math and English achievement than students from high income neighborhoods, but both groups rank higher achieving schools higher. The relationship between rank and distance is very similar to the results across baseline achievement levels. These differences suggest the importance of allowing for tastes for school achievement to differ by baseline achievement and income groups.

1.5.2 Model and Estimation Strategy

The comparison of the attributes of first choices relative to later choices provides rich information to identify how a student trades off these features of schools in her preferences. To quantify these tradeoffs, we work with a random utility model. Let i index students, j index programs, and let s_j denote the school housing program j. We model student i's indirect utility for program j using the following specification:

$$\begin{aligned} u_{ij} &= \delta_{s_j} + \sum_l \alpha^l z_i^l x_j^l + \gamma d_{ij} + \varepsilon_{ij}, \text{ with} \\ \delta_{s_j} &= x_{s_j}\beta + \xi_{s_j}, \end{aligned}$$

where z_i is the vector of student characteristics, x_j is a vector of program j's characteristics, d_{ij} is distance between student i's home address and the address of program j, ξ_{s_j} is a school-specific unobserved vertical characteristic, and ε_{ij} is an error term drawn from an extreme value type-I distribution with variance normalized without loss of generality to $\frac{\pi^2}{6}$. Since students rank programs, it is natural to specify utility in terms of programs. However, we write the mean utility δ_{s_j} only in terms of school attributes since a program's type and speciality are the only features that do not vary at the school level.

This demand specification allows us to exploit the unusually large number of observable school and student characteristics in our micro data; the bulk of our estimates come from models with 399 parameters, which include 284 school fixed effects. To take advantage of these characteristics, we estimate models with a large number of interactions, involving the interaction of each student characteristic with each school characteristics, and the logit functional form allows us to write the likelihood in closed form, sidestepping the need for numerical integration. The logit assumption also comes with the Independence of Irrelevant Alternatives (IIA) property, which implies that if the choice set changes, the relative likelihood of ranking any two schools does not change. Most of our counterfactual exercises compare two different allocations by aggregating utilities for two different assignments. We thus aim to best fit the utilities of students given their observable characteristics. Substitution between schools plays a less central role in this application because of the modest change in the set of school options between years.

In our preferred models, we do not explicitly include an outside option and instead normalize without loss of generality the value of δ for an arbitrarily chosen school to zero. This assumption is motivated by our primary interest in studying the allocation within inside options rather than substitution outside of the NYC public school system. As we describe below, we also find the behavioral implications of an outside option in rank order lists that do not rank all 12 alternatives unappealing. Indeed, the IIA property implies that our estimates do not change if we did observe the actual rank of the outside option.

The demand sample contains rankings of 69,582 participants in 2003-04 over 549 programs in 284 schools, representing a total of 547,011 school choices. We start by building a likelihood function assuming that choices are truthful before returning to examining this assumption in some detail. This assumption implies that student i ranks programs in order of the indirect utility she derives from the programs. We assume that all unranked schools are less preferred to all ranked schools. Let r_i be the rank order list submitted by student i in our demand sample, with $|r_i| \leq 12$ denoting the length of the agent's list and r_{ik} denoting the kth choice. Let $\theta = (\alpha, \gamma, \delta)$ denote the parameters of interest. Given θ and our logit assumption, the likelihood that student i submits rank order list r_i is:

$$\mathcal{L}\left(\theta|r_i, z_i, X, D_i\right) = \Pi_{k=1}^{|r_i|} P\left(u_{ir_{ik}} > \max_{j \in J \setminus \cup_{l<k} r_{il}} u_{ij} | \theta, z_i, x_j, D_i\right)$$

$$= \Pi_{k=1}^{|r_i|} \frac{\exp(\delta_{s_{r_{ik}}} + x_{ir_{ik}}\alpha + \gamma d_{is_{r_{ik}}})}{\sum_{j \in J \setminus \bigcup_{l<k}\{r_{il}\}} \exp(\delta_{s_j} + x_{ij}\alpha + \gamma d_{is_j})},$$

where $D_i = \{d_{ij}\}_j$ is the vector of distances from i's home to schools indexed by j. We estimate the parameters θ employing a two stage maximum likelihood. The first stage likelihood is written as a function of (δ, α, γ). β is estimated in the second stage from the equation $\delta_{s_j} = x_j\beta + \xi_{s_j}$ under the assumption $E\left[\xi|X\right] = 0$. Reported standard errors are from maximum likelihood. Since the number of students ranking programs is much larger than the number of programs, the estimation error in δ is negligible compared to the error in β due to sample variance in the set of observed schools.

1.6 Estimates and Model Fit

Our specifications follow other models of school demand and include average school test scores and racial attributes as school level attributes (Hastings et al., 2005). There are, of course, many features which make a school desirable that we do not observe. Therefore, it is worth noting that unlike in these earlier models, we include additive school fixed effects to proxy for school-level unobservables. As we have seen above, the distance to an assigned school is an important dimension on which choice changes, so we therefore always include a linear control for distance. We do not include more flexible controls for distance because it serves as our numeraire from which we measure the weight of other school characteristics and provides a metric for computing welfare in the counterfactuals. Moreover, the descriptive patterns shown in Table 1.5 suggest that the evaluation of distance does not differ across baseline achievement and income levels. In each specification, we also include three program type dummies and ten program speciality dummies.[5] The categorization of programs into specialities is described in the data appendix.

[5]The program type dummies are Unscreened, Screened, and Educational Option. The program specialty dummies are Arts, Humanities/Interdisciplinary, Business/Accounting, Math/Science, Career, Vocational, Government/Law, Other, Zoned, and General.

1.6.1 Parameter Estimates

Table 1.6 reports the parameter estimates for six specifications of our demand model. The first specification serves as our benchmark and involves 14 additional parameters (program specialty dummies, program type dummies, distance, and school characteristics) on top of the school fixed effects. It includes controls for five school characteristics: 9th grade school enrollment, percent white, attendance rate, percent free lunch, and the Regents math performance of students at the school. The second specification adds 34 parameters by allowing a more flexible relationship between all of the school characteristics with student's racial characteristics. The third specification interacts a school's enrollment, attendance rate, and math performance with baseline achievement measures of students. The next three specifications allows interactions between all available student demographics (gender, race, Special Ed, Limited English Proficient), measures of baseline achievement in math and english and our main school characteristics. We also include dummies for Spanish, Asian and Other Language Program, interacting these dummies with a student's LEP status and whether they are Hispanic or Asian. This model has 399 parameters. We report three variations of this specification, which use only the top ranked school, the top three ranked schools, and all ranked schools. Since these specifications include many interactions, we only report a subset of coefficients, deferring all of the estimates to Table A.6.

Across the first four specifications in Table 1.6 the coefficient on distance is -0.33, with little variation when we include more flexible student and school controls. The negative coefficient implies that everything else being equal, students prefer schools closer to home. The size of this coefficient relative to the other coefficients also implies that distance has an important weight on choices. The coefficients on the 9th grade enrollment and percentage white implies that students are willing to travel for bigger schools and for schools with higher white percentage. For example, a 15% increase in white percentage is equal to traveling an additional mile. Students also prefer schools with higher attendance rate, lower percentage of free lunch students, and higher Math performance of graduates. However, like enrollment and percent white, the estimates together with the variation in these attributes across schools would have a much smaller impact on choices than a 1 mile change in distance.

Table 1.6: Estimates from the Demand System

	School Char.	School Char. x Demographics	School Char. x Student Achievement	School Characteristics x Demographics School Characteristics x Achievement		
				All Choices	Top Choice Only	Top Three Choices
	(1)	(2)	(3)	(4)	(5)	(6)
A. Preference for Distance						
Distance	-0.318***	-0.313***	-0.316***	-0.310***	-0.387***	-0.353***
B. Preference for Schools						
9th Grade Enrollment						
Main effect	1.45E-04***	4.01E-04***	1.54E-04***	5.89E-04***	0.001***	9.85E-04***
Standardized Math Score			4.31E-05***	1.07E-05**	2.15E-05	7.62E-06
Standardized English Score			-5.50E-05***	-2.99E-05***	-8.92E-05***	-5.21E-05***
Free Lunch				-1.37E-04***	-2.94E-04***	-2.40E-04***
Special Ed				-7.05E-05***	-1.66E-04***	-1.22E-04***
Median Family Income				-6.50E-06***	-1.93E-05***	-1.93E-05***
Percent White						
Main effect	0.026***	0.040***	0.025***	0.032***	0.040***	0.033***
Asian		-0.016***		-0.015***	-0.027***	-0.018***
Black		-0.022***		-0.020***	-0.035***	-0.024***
Hispanic		-0.011***		-0.009***	-0.016***	-0.010***
Attendance Rate						
Main effect	0.055***	0.103***	0.062***	0.075***	0.131***	0.117***
Standardized Math Score			0.012***	0.010***	0.017***	0.015***
Standardized English Score			0.014***	0.014***	0.020***	0.021***
Free Lunch				-3.48E-04	-0.008***	-0.005***
Median Family Income				0.006***	0.008***	0.007***

Table 1.6: Estimates from the Demand System (Cont'd)

	School Char.	School Char. x Demographics	School Char. x Student Achievement	School Characteristics x Demographics School Characteristics x Achievement		
				All Choices	Top Choice Only	Top Three Choices
	(1)	(2)	(3)	(4)	(5)	(6)
Percent Free Lunch						
Main effect	-0.002***	-0.007***	-0.003***	-0.004***	0.005***	-0.003**
Asian		3.96E-04		-8.29E-04	-0.004***	-8.80E-04
Black		0.002***		0.001**	-0.008***	-0.002***
Hispanic		0.010***		0.009***	0.007***	0.010***
Free Lunch				1.49E-04	-0.002**	-0.001***
Median Family Income				-7.22E-04***	-0.001***	-7.36E-04***
Math Performance of Graduates						
Main effect	0.008***	0.007***	0.004***	2.68E-04	0.000765	-0.000866
Standardized Math Score			0.008***	0.007***	0.012***	0.009***
Standardized English Score			0.005***	0.005***	0.006***	0.005***
Free Lunch				-0.003***	-0.005***	-0.005***
Special Ed				-0.003***	-1.07E-04	-1.00E-03
Median Family Income				1.19E-04*	-7.12E-04***	-2.42E-04**
Limited English Proficiency				-0.002***	-0.003*	-0.005***

Table 1.6: Estimates from the Demand System (Cont'd)

	School Char.	School Char. x Demographics	School Char. x Student Achievement	School Characteristics x Demographics School Characteristics x Achievement		
				All Choices	Top Choice Only	Top Three Choices
	(1)	(2)	(3)	(4)	(5)	(6)
Spanish Language Program						
Limited English Proficiency (LEP)				5.147***	5.755***	5.472***
LEP Hispanic Student				-3.142***	-3.817***	-3.147***
Asian Language Program						
Limited English Proficiency (LEP)				3.687***	4.253***	4.203***
LEP Asian Student				-2.131***	-3.299***	-2.717***
Other Language Program						
Limited English Proficiency (LEP)				2.289***	2.947***	2.870***
Program Type Dummies	X	X	X	X	X	X
Program Specialty Dummies	X	X	X	X	X	X
C. Model Fit						
Number of ranks	547,011	547,011	547,011	547,011	69,582	196,408
Number of parameters	248	272	280	349	349	349
Pseudo R-squared	0.85	0.89	0.89	0.92	0.96	0.95
R-squared of 2nd-Stage δ Regression	0.43	0.68	0.42	0.53	0.58	0.60

Notes: Coefficients from demand systems with 69,907 students and submitted ranks over 497 program choices in 234 schools. Except columns (4) and (5), all columns use all submitted ranks. Column (1) contains no interactions between student and school characteristics. Column (2) contains interactions of race dummies with all school characteristics. Column (3) contains interactions of Math and English score with all school characteristics. Columns (4-6) contains interactions of gender, race, achievement, special needs and income with all school characteritics. Number of parameters correspond to parameters outside of delta. R-squared of 2nd stage delta regression projects mean utility on school attributes. Significance at 90% (*), 95% (**) and 99% (***) confidence.

Many of the interaction terms are significantly estimated and have the expected signs. For instance, students with high baseline math scores tend to prefer schools with high Math performance more than those with low baseline scores in column (3). The interaction coefficients in column (6) predict significant heterogeneity across sub-populations. Looking at the full set of estimates in Table A.6, black, Hispanic and Special Ed. students are less willing to travel for bigger schools. Willingness to travel also decreases with family income, as indicated by the negative estimate on the interaction term with median family income. Also, non-white students are less likely to rank schools with a higher White percentage. Moreover, students with higher Math or English scores or from wealthier families prefer high performing math schools, whereas Special Ed and Limited English Proficient students prefer them less than the average student. Though many of these interaction terms are significantly estimated, they do not lead to a large improvement the our psuedo-R^2 which is about 0.85, suggesting that even after controlling for distance, the school and program fixed effects, student preferences remain substantially heterogeneous. This fact is also apparent looking at our measure of variance across students in the average utility for a program. There almost no change in this measure across the first four columns.

1.6.2 Assessing the Behavioral Assumptions

Treating all preferences as truthful, as in the estimates in the first four columns of Table 1.6, is a natural starting point and generates sensible patterns. It is a benchmark because of the straightforward incentive properties of the mechanism and the advice that the NYC DOE provides in the 2003-04 High School Directory and elsewhere. This includes statement that participants should "rank your twelve selections in order of your true preferences" with the knowledge that "schools will no longer know your rankings." Nonetheless, truthful behavior is a strong assumption worth investigating for a number of reasons.

A first issue is whether students are deliberately ranking schools, given that there may be substantial frictions involved in evaluating schools. For instance, Kling et al. (2012) document "comparison frictions" in the evaluation of different Medicare Part D prescription drug plans. The choice between a vast number of school programs may be daunting and as a result, some may simply randomly list choices, especially further down their rank order list. Of course, if preferences were generated in this way by a large fraction of participants, then we'd expect that most of our point estimates to be imprecisely estimated or have unintuitive patterns, but they do not. Another way to see that choices matter for participants is to examine enrollment decisions by choice. Table A.5 reports the assignment and enrollment decisions for students who are assigned in the main round. The table shows that 92.7% of students enroll in their assigned choice, and this number varies from 88.4% to 94.5% depending on which choice a stuent obtains. Interestingly, the highest take-up is for students who receive later choices, while the fraction of students who exit for private school is highest among students who obtain one of their top three choices. This fact suggests that either families are indifferent between later choices and simply enroll where they obtain an offer, families accepted to their top choice may have good outside options, or families have deliberately investigated later choices and are happy to enroll in a lower ranked school.

If families are more uncertain about lower ranked choices, then using the ranking information contained in all of these choices may provide a misleading picture of preference parameters. The last two columns of Table 1.6 reports demand estimates using only the top and top three ranked schools. These smaller samples use only 13% and 36% of our the submitted choices, so they throw away much of the information in our data set. Despite this, the coefficient on distance is similar across the three models. Many of the interaction terms have similar signs to column (4), but as expected, they are estimated less precisely using a fewer number of choices. On balance, the pattern of estimates shows remarkable consistency.

The second issue with the assumption of truthful preferences is that students can rank at most 12 programs on school applications. When a student is interested in more than twelve schools, she has to carefully reduce her choice set down to at most twelve schools. If a student is only interested in 11 or fewer schools, this constraint in principle should not influence her ranking behavior. It is a dominant strategy to add an acceptable school to the list as long as there is room for additional schools in the application form. However, 22.6% of students in our demand sample rank 12 schools. These students are likely to drop highly sought-after schools from top of their choice lists.

To probe how this issue impacts our preference estimates, in Table 1.7, we report estimates where we drop students ranking 12 choices from the sample. Despite the change in the composition of students, there is very little change in our estimated preference parameters. The coefficients on distance and school main effects are nearly identical with the full model, implying that the 12 choice constraint will not significantly impact our conclusions. The other variation we report only uses information about the choice among ranked alternatives. This model assumes that when a student does not rank a school, it is missing at random: the school could either be ranked above an applicant's top choice or below it, while our models so far treat ranked schools as more preferred than unranked schools. Perhaps unsurprisingly, the estimates from this model, in column (3), are considerably different than our other specifications. In particular, the importance of distance and school main effects are not comparable to the other specifications. The psuedo R^2 is only 0.24 and the model fit is worse.

A third threat to the truth-telling assumption is that parents rank schools using heuristics garnered from the previous system. Despite the theoretical basis and advice, parents might still deviate from truth-telling for reasons related to misinformation or lack of information about the new mechanism. Table A.4 shows that students are more likely to be assigned their last choice than be assigned the one just above the last choice in their rank lists. This pattern may be caused by strategic behavior if students apply to schools that they like and, as a safety option, put in the last choice a school with which they have a higher chance of admissions. However, this pattern may also be fully consistent with truth-telling. For example, students usually obtain borough priority or zone priority in their neighborhood schools, which significantly improves their likelihood of being assigned to these schools in case they are rejected by their higher choices and apply to them within the algorithm. If students consider applying and commuting to schools further away from their neighborhood for reasons like Math and English achievement, they may as well stop ranking schools below their neighborhood schools once such considerations no longer justify the cost of commute. This preference pattern would produce the observed assignment pattern in the data. Column (4) in Table 1.7 presents estimates which drop the last choice of applicants; the coefficients on distance and school effects are nearly identical to column (1), implying that the possibility of strategic safety schools seems unlikely to alter our conclusions.

Finally, our approach avoids use of rank order lists to identify the value of the outside

Table 1.7: Demand Estimates from Alternative Specifications

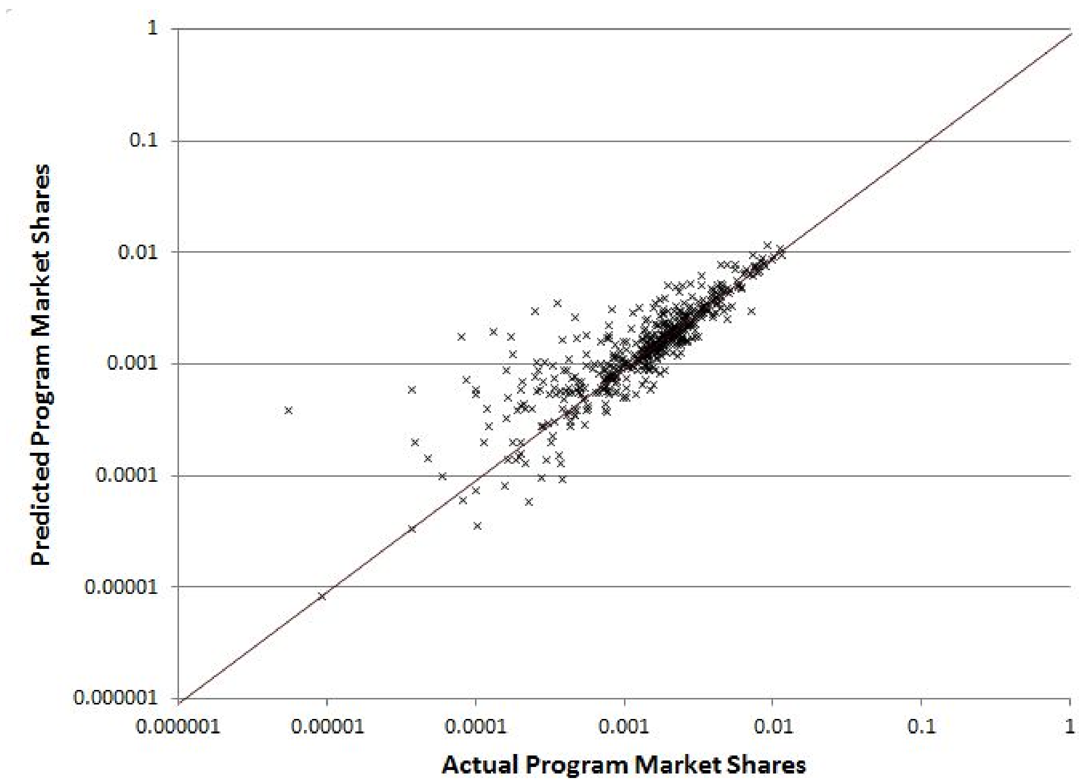
	All Choices (1)	Drop lists wth 12 choices (2)	Dropping last Choice (3)	Ranked Only (4)	With Outside Option (5)
A. Preference for Distance					
Distance	-0.310***	-0.317***	-0.302***	-0.024***	-0.304***
B. Preference for Schools - main effects					
Standardized Math Score					
9th Grade Enrollment	5.89E-04***	6.68E-04***	6.12E-04***	0.000418	4.43E-04***
Percent White	0.032***	0.032***	0.029***	1.00E-03	0.030***
Attendance Rate	0.075***	0.082***	0.087***	6.10E-02	0.053***
Percent Free Lunch	-0.004***	-0.004***	-0.004***	5.00E-03	-0.005***
Math Performance of Graduates	2.68E-04	-1.00E-03	0.002***	3.00E-03	0.002***
Program Type Dummies	X	X	X	X	X
Program Specialty Dummies	X	X	X	X	X
C. Model Fit					
Number of ranks	547,011	358,119	358,120	358,121	358,122
Log-Likelihood	-2,690,008	-1,818,749	-2,431,624	-753,816	-2,876,073
Number of parameters	349	349	349	349	350
Variance within students	5.62	5.88	5.50	0.44	5.45
Variance between students	12.26	13.26	12.97	1.99	5.00
Pseudo $R\hat{2}$	0.92	0.92	0.92	0.60	0.86
R-squared of 2nd-Stage Delta Regression	0.53	0.56	0.51	0.15	0.47

Notes: Coefficients from demand systems with variations in the sample of students and alternative behavioral assumptions. All estimates have same controls as in column (6) of Table 1.6, which includes interactions with student racial and achievement attributes and school characteristics. The first column treats all choices as truthful, the second column drops students who have ranked 12 schools from the sample, the third column uses information about relative comparisons among listed alternatives, assuming that unlisted alternatives are missing at random, and the fourth column drops the last ranked school from each students list, and Number of parameters correspond to parameters outside of delta. Significance at 90% (*), 95% (**) and 99% (***) confidence.

option relative to NYC public schools. The primary reason is that it that our counterfactuals exercises focus on the re-allocation of inside goods. However, one may be worried about bias in our taste estimates from this assumption. If ranking additional acceptable schools is cognitively costless, it may be appropriate to assume that unranked schools are worse than the outside option. There are two indications that this is not true: the median student ranks 9 out of 649 programs and 59% of students who are unassigned in the main round finally enroll in a program that they did rank in the main round. Ignoring the outside option does not yield biased estimates if the value of outside option is not related to preferences for inside options (IIA) or if number of ranked schools is unrelated to the value of outside option. These assumptions may not be satisfied, so we report estimates with an outside option in the last column of Table 1.7, where we normalize the value of the outside option to zero. Our IIA assumption implies that estimates should not be sensitive to our specification of the outside option that is assumed to be preferable to all unranked schools, and our estimates are close to those in the first column. However, all of the school-specific δ parameters are less than 0, suggesting that students would prefer the outside option (i.e. not a NYC public school), which contradicts with the high take-up of assigned school among our students.

1.6.3 Model Fit

Before reporting our counterfactuals, we now turn to model fit by first reporting on measures of in-sample fit. Figure 1.3 shows a plot of the predicted market shares of each program against the program's actual market share on a log scale. Market share is the fraction of students who rank a program anywhere on their rank order list. We compute predicted market shares by drawing the top 12 programs from the estimated preference distribution 100 times for each student, and taking the average for each program. The figure shows a strong correlation between the two measures, an encouraging phenomenon likely driven by the presence of school fixed effects and the large number of student and school interactions. However, this relationship is not mechanical since the parameters are not estimated by inverting the fraction of students ranking a program first (Berry, 1994). The single set of school fixed effects used to explain multiple rank could have resulted in a poor fit.



Market share is the fraction of applicants ranking pra program anywhere on rank order list. Predicted program market shares based on preference estimates from preferred specification in Table 6, using 100 draws of out estimates and assuming students rank their top 12 programs. Each point corresponds to the predicted and actual market share.

Figure 1.3: Log-Log Plot of Predicted vs. Actual Program Market Shares

Table 1.8: In-Sample Model Fit

		Overall	Ranked Choice											
			1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th	12th
Distance	Observed	12.3	4.4	4.8	5.1	5.2	5.4	5.5	5.6	5.6	5.6	5.6	5.4	5.1
	Simulated		5.1	5.2	5.3	5.3	5.4	5.4	5.5	5.5	5.6	5.6	5.7	5.7
Attendance Rate	Observed	85.8	87.2	86.7	86.4	86.1	85.9	85.7	85.5	85.5	85.2	85.0	84.8	84.5
	Simulated		86.3	86.2	86.2	86.1	86.1	86.1	86.0	86.0	86.0	86.0	86.0	86.0
Percent Free Lunch	Observed	62.9	51.4	53.4	54.5	56.2	57.4	58.7	59.8	60.6	61.3	61.9	62.7	63.1
	Simulated		53.5	53.7	54.0	54.2	54.4	54.6	54.7	54.9	55.0	55.2	55.3	55.4
Fraction Inexperienced Teachers	Observed	55.9	43.0	43.8	45.0	46.2	47.1	48.1	48.7	49.5	49.5	50.2	49.9	49.9
	Simulated		44.9	44.9	45.0	45.1	45.1	45.2	45.3	45.3	45.4	45.4	45.5	45.5
High Math Achievement	Observed	10.4	16.7	15.3	14.7	13.9	13.4	12.8	12.4	11.9	11.5	11.1	10.8	10.4
	Simulated		14.9	15.0	14.7	14.5	14.3	14.2	14.1	14.0	14.0	13.9	13.9	13.8
High English Achievement	Observed	19.3	27.8	25.6	24.3	23.2	22.2	21.4	20.8	20.5	19.9	19.4	19.1	18.6
	Simulated		24.8	24.8	24.5	24.3	24.0	23.9	23.7	23.6	23.5	23.5	23.4	23.3
Percent Attending 4yr College	Observed	47.2	53.7	54.1	52.8	51.7	50.9	50.1	49.3	49.3	49.0	48.3	47.6	47.5
	Simulated		52.8	53.1	53.1	53.0	52.9	52.7	52.6	52.5	52.4	52.3	52.2	52.1
Percent White	Observed	10.8	19.1	16.7	15.7	14.4	13.3	12.2	11.2	10.8	10.4	9.8	9.4	9.0
	Simulated		17.4	17.1	16.7	16.5	16.2	16.0	15.9	15.7	15.6	15.4	15.3	15.1
Percent Black	Observed	38.1	32.6	34.3	35.1	36.0	36.7	37.5	38.2	38.4	38.7	38.9	38.9	38.1
	Simulated		34.5	34.7	34.9	35.1	35.2	35.4	35.5	35.6	35.7	35.8	35.9	36.0
Percent Hispanic	Observed	42.1	34.4	35.2	35.9	37.0	37.9	38.9	39.6	40.2	40.8	41.5	42.2	43.5
	Simulated		35.1	35.2	35.4	35.5	35.5	35.7	35.7	35.8	35.8	35.9	36.0	36.0
Percent Asian	Observed	9.0	13.8	13.7	13.3	12.7	12.2	11.5	11.0	10.6	10.1	9.7	9.5	9.4
	Simulated		13.0	13.0	13.0	13.0	13.0	13.0	13.0	12.9	12.9	12.9	12.9	12.9

Notes: Table reports the observed characteristics of each choice compared to those from our estimated preference distribution corresponding to the preferred model in Table 1.6, Column (4). The first column replicates the average school summary statistics shown in Table 1.2, where distance corresponds to the average distance to a school.

Another way to gauge within-sample fit is to see what our estimates imply for the aggregate patterns of choice shown in Table 1.5. In Table 1.8, we use the estimates to compute the distribution of preferences and aggregate them into cells corresponding to those reported in Table 1.5. Comparing the ranked program means correspond to the overall distribution of school characteristics, shown in the first column, the preference estimates capture many features of the distribution of preferences, though the estimated tradeoffs in school attributes going down rank order lists is sometimes not as steep as what is observed. For instance, the average high math achievement of a first choice is 14.6, while it drops to 13.8 for last choices. In the observed data, the corresponding range is 16.8 to 10.5. Relative to the average attributes of schools, the model fit is much closer to the actual ranked distribution. The average distance to a high school in New York is 12.3 miles away from home, yet the top ranked school is 4.5 miles away. The overall percent free lunch in New York's schools is 62.9, but the average free lunch percentage for the first choice is 51.4 and in our preferred estimate, it is 54.0. Moreover, the average percent white in New York's schools is 10.8, but first choices are 19.2 white, while we predict them to be 17.0. Each of our preference predictions are closer to the actual declared properties of ranked schools, relative to the aggregate distribution.

The attributes of schools actually ranked and what we predict are close, especially for choices 3-10. There are more significant differences between the prediction and actual choices for the first two choices. For instance, we predict that the average first choice is 5.2 miles away, while it is 4.5 miles away in the data. It is worth noting that a student's closest school is 0.86 miles away, so our predictions are far better than a behavioral model where students simply rank their closest school first. The drop in predictive accuracy for the last three choices is probably driven by changes in the composition of students who rank this many choices. Nonetheless, the model appears to capture the trend in preferences going down rank order lists.

We next relate our estimated taste parameters to information that we have not used for estimating the distribution of preferences, an out-of-sample test. In Table 1.9, we regress the earlier measures of exit and noncompliance on our estimates of the observable components of student i's utility for his assignment. Let y_i indicate whether a student exits NYC public schools or enrolls in a school other than her assigned one. We report estimates of

Table 1.9: Out of Sample Fit for Exit and Noncompliance

	Centralized Mechanism		Decentralized Mechanism	
	(1)	(2)	(3)	(4)
	A. Exit from NYC Public Schools			
Utility from Assignment (in mi)	-0.0021***		-0.0029***	
	(0.0002)		(0.0002)	
Distance from Assignment		-0.0002		-0.0026***
		(0.0003)		(0.0003)
Observations	66,921	66,921	70,358	70,358
R-squared	0.0687	0.0671	0.0926	0.0916
	B. Noncompliance with Assignment			
Utility from Assignment (in mi)	-0.0140***		-0.0161***	
	(0.0003)		(0.0004)	
Distance from Assignment		0.0121***		0.0010**
		(0.0004)		(0.0005)
Observations	62,656	62,656	64,349	64,349
R-squared	0.0621	0.0404	0.0622	0.0337

Notes: Linear probability models. Exit is enrollment in a school outside the NYC Public School System conditional on assignment. Enrollment in unassigned school is conditional on enrollment in a NYC Public School. Decentralized mechanism refers to 2002-2003 mechanism; centralized mechanism refers to the deferred acceptance mechanism adopted in 2003-2004. Utility estimates from the preferred specification, projected on observables alone. All columns include student borough, race, limited English proficiency, special ed, no Math score, no English score and SHSAT taker dummies. All columns also include the median family income from the student's census block group and standardized Math and English scores. Significance at 90% (*), 95% (**) and 99% (***) confidence.

the following equation:

$$y_i = \lambda \hat{u}_{ia} + \gamma d_{ia} + z_i \beta + \epsilon_i,$$

where $\hat{u}_{ia}$ is our estimate of student i's utility for his assigned program a in miles, d_{ia} is the distance to assigned school, and z_i represents the student demographic and baseline achievement measures. Since distance is a natural component of the student's preferences and a key aspect of student preferences, we are interested in whether our imputed utility provides additional information beyond distance for the exit or noncompliance decision, conditional on student demographics.

The estimates in column (3) show that if students obtain greater utility for their assignment, they are less likely to exit in the centralized mechanism. This pattern is reassuring since the demand estimates come from the main round of the centralized mechanism, but we have not used information on the exit decision in estimation. Distance to school has no additional predictive power for the exit decision beyond the controls for student demographics in column (4). Noncompliance is also negatively related to the utility from assignment in Panel B, while distance exhibits the opposite pattern. The magnitude of an increase distance is smaller than the magnitude of an decrease in estimated utility (in distance units) on exit.

An even more demanding comparison involves considering what our estimates imply for data from the decentralized mechanism. This comparison uses data on preferences from a different year to look at an out-of-sample decision in that year. It is reassuring that the relationship between utility, exit and noncompliance is similar to that from the those reported using data from the centralized mechanism. Just as with the centralized mechanism, student exit is negatively related to a student's utility from their assigned school, and the utility measure has a larger weight than distance. The magnitude of the estimate for noncompliance is nearly identical to that in the new mechanism. This out-of-sample comparison suggests that our preference estimates may be suitable for understanding the utility associated with the assignments in the old mechanism.

1.7 Comparing Mechanisms

1.7.1 Measuring Welfare

The preference estimates in the last section allow us to compare the assignments produced by the decentralized and centralized mechanisms. Consider a group of students with attribute G and a matching μ, which specifies the program for each student as $\mu(i)$, where if student i is unassigned $\mu(i) = i$. Define the average welfare as a function of parameters θ as

$$W_G^{\mu}(\theta) = \frac{1}{|G|} \sum_{i \in G} u_{i\mu(i)}(\theta),$$

which is the per student average of the utilitarian social welfare criterion for the group G. For two different matchings, μ and μ', corresponding set of students in demographic group $G(\mu)$ and $G(\mu')$, the estimate of the welfare difference between the two matchings is given by:

$$W_G^{\mu}(\theta) - W_G^{\mu'}(\theta) = \frac{1}{|G(\mu)|} \sum_{i \in G(\mu)} u_{i\mu(i)}(\theta) - \frac{1}{|G(\mu')|} \sum_{i \in G(\mu')} u_{i\mu'(i)}(\theta).$$

Notice that the welfare difference depends only on the utility differences between the various inside options at any parameter θ. This comparison may understate the total impact of the new assignment system because it is only an intensive margin calculation and does not include the difference in exit rates that we described earlier.

To operationalize this formula, we have to confront three issues. First, there is a different set of students for the two different matchings, so for this comparison to be valid we must assume that the preference estimates for students in the new mechanism are relevant for those in the old mechanism. Second, for each student we compute the implied utility given the observed characteristics of the student, but do not know the student's unobserved component given by ϵ_{ij}. We focus only on the observed component, ignoring the contribution of ϵ_{ij}. Therefore, our analysis is about the particular assignment produced by the mechanism. The reason we have to ignore the unobserved component of student tastes is that we have no way to simulate the outcomes of the old mechanism given the student preferences. Finally, our comparisons may understate the welfare gains from the new mechanism given the sizable number of students who eventually switched schools in the old mechanism. It seems reasonable that students do not comply only if they prefer another school over what the assignment mechanism produced.

Let $\hat{\theta}$ denote the parameter estimates, and define the expected utility for student i with characteristics z_i and distance vector D_i with school attributes X from assignment μ as

$$\hat{u}_{i\mu(i)} = \mathbb{E}[u_{i\mu(i)}|\hat{\theta}, z_i, X, D_i] = \hat{\xi}_{\mu(i)} + x_{\mu(i)}\hat{\beta} + \sum_l \hat{\alpha}^l z_i^l x_{\mu(i)}^l + \hat{\gamma} d_{i\mu(i)}.$$

No unassigned students are in the welfare samples, so it is not necessary to impute a value to these students. Likewise, if a student is assigned to a school that is closed in 2002-03, we set $\hat{\xi}_{\mu(i)} = 0$. Therefore, for a group of students G, the contrasts between the two mechanisms are:

$$\hat{W}_G^{\mu} - \hat{W}_G^{\mu'} = \frac{1}{|G(\mu)|} \sum_{i \in G(\mu)} \hat{u}_{i\mu(i)} - \frac{1}{|G(\mu')|} \sum_{i \in G(\mu')} \hat{u}_{i\mu'(i)}.$$

This formula illustrates that only relative comparisons are meaningful, so when we compare the decentralized and centralized mechanism, we normalize the utilities so that the mean utility for all students is zero.

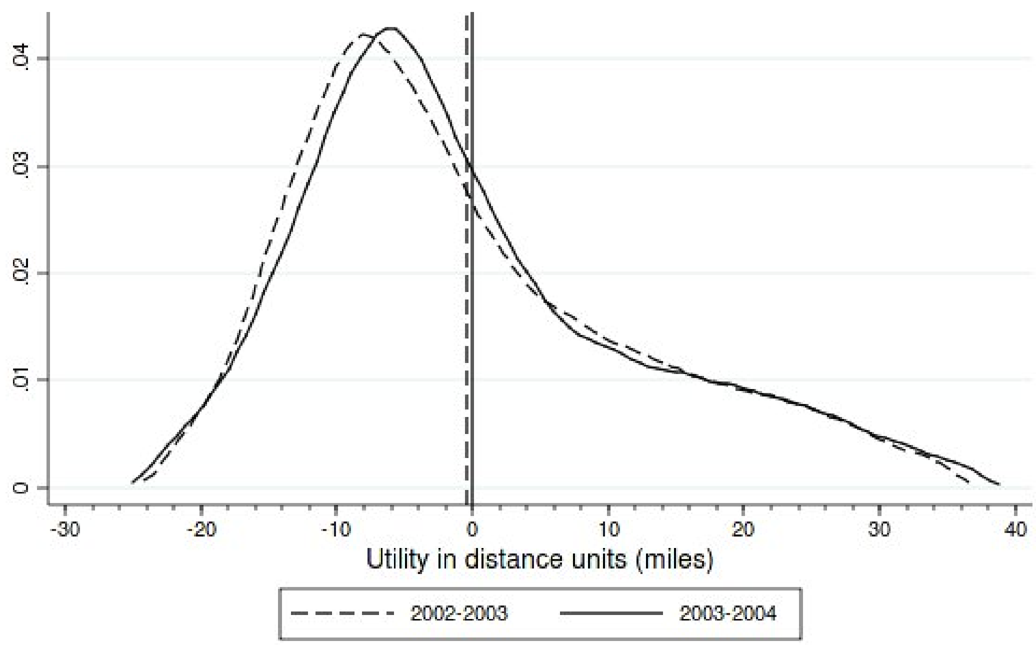
1.7.2 Welfare Impact of the Centralized Mechanism

On average students benefit by a utility equivalent of 0.25 miles from the new mechanism. The overall distribution of student utilities (in distance units) is shown in Figure 1.4, which stands in contrast to the increase in distance to assigned school shown in Figure 1.2. The reason for the difference is that students are more likely to prefer their assigned school in the new system. We report the average utility from the two assignments for different student groups in Table 1.10. All measures are normalized so that the mean utility in the decentralized mechanism is equal to zero. Asians and whites tend to travel less than blacks and Hispanics in the decentralized mechanism. Moreover, students from Staten Island and Brooklyn travel less, as do low baseline students measured by Math achievement. Many of these patterns also remain true within the assignment produced by the centralized mechanism, and may simply reflect cross-sectional differences in access to desirable school options for these students.

Table 1.10: Welfare Comparison between Centralized and Decentralized Mechanisms

	Utility from Assigned Program				Change in Utility			
	Decentralized Mechanism		Centralized Mechanism		Mean		Median	
	Mean	Median	Mean	Median	Overall	Due to Distance	Overall	Due to Distance
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
All Students	0.00	-3.15	0.21	-2.91	0.21	-0.57	0.24	-0.68
Female	0.44	-2.78	0.67	-2.50	0.24	-0.58	0.28	-0.67
Asian	16.85	15.77	18.77	18.16	1.92	-0.55	2.39	-0.44
Black	-7.05	-7.50	-6.86	-6.93	0.19	-0.65	0.56	-0.77
Hispanic	-5.13	-5.75	-4.61	-5.01	0.52	-0.48	0.73	-0.69
White	17.69	17.57	18.60	18.58	0.92	-0.50	1.01	-0.32
Bronx	-7.30	-8.31	-7.46	-8.00	-0.16	-0.71	0.31	-0.79
Brooklyn	0.24	-2.69	0.37	-2.49	0.12	-0.49	0.19	-0.58
Queens	3.04	0.83	3.30	0.81	0.27	-1.05	-0.02	-1.30
Manhattan	-2.04	-5.43	-1.55	-4.95	0.49	0.10	0.47	0.17
Staten Island	15.41	17.46	16.57	18.16	1.16	-0.21	0.70	-0.06
High Math Achievers	12.38	10.60	13.05	10.92	0.66	-0.45	0.32	-0.52
Mid Math Achievers	-0.87	-3.20	-0.62	-3.00	0.25	-0.58	0.20	-0.70
Low Math Achievers	-10.06	-10.84	-9.77	-10.60	0.29	-0.41	0.24	-0.55
No Math Score	-3.52	-6.71	-5.45	-7.85	-1.92	-0.96	-1.14	-0.92
Special Education	-9.73	-10.85	-11.00	-12.06	-1.27	-0.62	-1.21	-0.75
Limited English Proficiency	-3.61	-6.07	-2.68	-5.15	0.93	-0.39	0.92	-0.55
SHSAT Takers	7.66	4.96	8.43	5.49	0.77	-0.48	0.53	-0.64
Free Lunch	-1.67	-4.54	-0.92	-3.81	0.75	-0.50	0.72	-0.59
Low Income Census Tract	-7.10	-8.32	-6.86	-7.71	0.23	-0.44	0.61	-0.56

Notes: Utilities projected on the observable characteristics. All utilities are monetized in distance units (miles) and normalized to zero at the mean utility to all students in the Decentralized Mechanism. Statistics from Mechanism Comparison Sample described in Table 1. Students with a score under the 25th percentile (over the 75th percentile) in the middle school match examination are categorized as having a Low Baseline Score (High Baseline Score).



Distribution of Utility (measured in distance units) plotted with mean utility in 2003-2004 normalized to 0. Average difference corresponds to 0.21 miles. Top and bottom 1% are not shown in figure. Line fit from Gaussian kernel with optimal bandwidth.

Figure 1.4: Student Welfare from New Mechanism

A more interesting comparison is to look across mechanisms. The average student in the new mechanism obtains a school that is the equivalent of 0.25 miles more preferred than their assignment in the old mechanism. Since the average student is going to a school which is 0.7 miles further from home (shown in column (6)), the additional 0.95 miles are due to a better match between a student's preferences and the school program. The assignment of blacks and Hispanics is worth 0.11 and 0.35 miles more than their decentralized assignment, while Asian and whites realize an increase of 0.87 and 0.74 miles in utility, respectively. A large part of this differential gain comes from differentials in distance: white students are able to attend schools 0.55 miles further, while black students are matched to schools 0.8 miles further.

As with different racial groups, students from each of the boroughs prefer the new assignment, with students in Staten Island gaining the most, and students from the Queens and Manhattan gaining the least. The gains are similar across Math achievement levels, Special Education Status, Limited English proficient, and measures of income. Although not shown, we have repeated these comparisons using alternative preference estimates from Table 1.6, and the pattern is qualitatively similar.

One issue with these comparisons is that the set of school options changed in 2003-04 with the arrival of new small schools. Since the new schools are unknown quantities, they are not ranked by many students and hence students who are assigned there obtain low utility. However, this seems more likely driven by inadequate information about these schools since they were not listed in the high school directory. Therefore, it's worth emphasizing that the sample in Table 1.10 does not include students who are assigned to schools announced after the printing of the high school directory.

1.7.3 Design Tradeoffs

Many features of the centralized mechanism were intended to address issues created by the old mechanism. Among the difficulties of the old mechanism were congestion, student's incentives to strategically rank schools, and school's incentives to conceal capacity and to distort student rankings by giving priority to student's first choices. Despite some exceptions, the new mechanism attempts to address all of these issues. But is the allocation produced by the new mechanism the best possible one, or would alternative assignment mechanisms improve overall student welfare? Of course, if there were more high desirable schools, then student welfare seems likely to increase regardless of the assignment mechanism. We therefore focus attention on changes to the assignment mechanism holding fixed the set of schools.

To measure the maximum possible student welfare, we use the cardinalization of utility implied by our demand model. We compute the assignment that maximizes the sum of student utility subject the feasibility constraints of the assignment for the demand sample. This program corresponds to the utilitarian optimal assignment, where we ignore the priorities of students at schools and solve for the best student feasible school following (Shapley and Shubik, 1971). Since we can recreate the assignments in the new mechanism and for other alternatives, we explicitly account for unobserved components of student preferences unlike in our last counterfactual. We first draw student preferences randomly according to the estimated distribution, now explicitly accounting for ϵ_{ij} to capture the unobserved component of student tastes, taking 1,000 draws from the distribution conditional on the observed ranks. We then solve the optimal assignment problem for each of the 1,000 draws.

With this aggregate social-welfare maximizing level of utility as our benchmark, the first question we ask is how does the allocation produced by the current mechanism compare to the maximum possible attainable total utility for students. We compute assignments from the student-proposing deferred acceptance algorithm using stated preferences and a single tie-breaking rule 1,000 times. To avoid imputing a welfare to being unassigned, if a student in the demand sample is left unassigned, we use their preferences to mimic NYC's supplementary round, by assigning them under a serial dictatorship according to the random ordering of students. Student preferences for this round use the simulated ranks from the estimated model. Following the computation for the optimal assignment, the welfare calculation from this assignment accounts for the effect of unobservable taste components ϵ_{ij}. The first column of Table 1.11 shows that the allocation produced by the new mechanism is 2.74 miles from the maximum achievable without any school-side constraints other than capacity. Relative to the difference between the decentralized and centralized assignment mechanism of 0.25 miles, this contrast implies about coordinated assignment accounts for 10% of the maximal possible improvement for students within the constraints of the existing algorithm.

Table 1.11: Welfare Comparison between Mechanisms Relative to Utilitarian Optimal Assignment

	Deferred Acceptance (Single Tiebreaking) (1)	Student-Optimal Matching (2)	Pareto Efficient Matching (3)
Utility in miles	-2.74	-2.63	-2.13
Reallocations relative to DA		2343.72	10881.17

Notes: Means. Results from 1,000 simulations of deferred acceptance with single tie-breaking using the Demand Sample. Utilities in terms of distance numeraire and include the effect of unobservables. Mean utility from the utilitarian optimal normalized to zero.

The utilitarian optimal assignment serves as an idealized benchmark, but is unlikely to be achieved given that it requires a cardinal mechanism and completely ignores the preferences of schools. There are 208 screened programs in New York City in 2003-04, so implementing this allocation would involve overriding the preferences of many of these programs. It is also possible that students would express different rank orderings if the mechanism were announced to produce this outcome. Therefore, another alternative we consider does not completely abandon school priorities. The student-proposing deferred acceptance (DA) algorithm produces a stable matching, which cannot be blocked by a student and school pair. The presence of ties in school's ranking of students means that DA does not produce a student-optimal stable matching, even though it would if all school preferences are strict. However, a number of authors have pointed out that deferred acceptance cannot be improved upon without sacrificing strategy-proofness for students (Abdulkadiroğlu et al., 2009; Erdil and Ergin, 2008; Kesten, 2010; Kesten and Kurino, 2012).

We quantify the cost of providing straight-forward incentives for students by computing a student-optimal stable assignment that improves the DA assignment by assigning students higher in their choice lists. Such an assignment can be computed by the stable improvement cycles (SIC) algorithm developed by Erdil and Ergin (2008), which iteratively finds Pareto improving swaps for students, while still respecting the stability requirement for underlying weak priority ordering of schools. On average, 2,344 students in the demand sample can obtain a better assignment in a student-optimal stable matching, and this corresponds to an average difference of 2.63 miles of equivalent utility with the utilitarian optimal assignment. The benefit of a student-optimal stable matching corresponds to a difference of 0.11 miles relative to the current student-proposing deferred acceptance algorithm. The cost of this change is that the underlying mechanism is not based on a strategy-proof algorithm.[6]

One step in further relaxing the constraints imposed by the algorithm is to abandon stability. While no stable mechanism eliminates strategic maneuvers by schools, if a mechanism does not produce a stable outcome, it is possible that schools benefit by offering students seats outside of the assignment process. Stability may therefore be motivated as an incentive constraint on the assignment mechanism to deal with strategic schools.[7] Despite the SIC-outcome being student-optimal stable, it is not Pareto efficient for students. We quantify the cost of school incentives by computing a Pareto efficient assignment that improves up on the student-optimal stable assignment by assigning students higher in their choice lists. A Pareto efficient assignment can be found by transferring students from their assigned schools to their higher ranked choices via the Gale's Top Trading Cycles algorithm (Shapley and Scarf, 1974).

Consequently, we calculate a Pareto efficient matching which dominates each student-optimal stable matching we simulate and estimate students' utilities from this Pareto efficient matching in column (3) of Table 1.11. On average, 10,881 students obtain a more preferred assignment at a Pareto efficient matching, which is 2.13 miles away for each student from the utilitarian optimal assignment. Relative to the current mechanism, the cost of limiting the scope for strategizing by schools (by imposing stability) is 0.61 miles per student. Both the "cost of strategy-proofness for students" and the "cost of stability" are comparable to the "benefit of coordination" from the centralized mechanism. Even if it were possible to implement a student-optimal stable matching or Pareto efficient matching using some mechanism, the benefits are similar to those implemented by NYC's switch to

[6] Azevedo and Leshno (2011) provide an example where the equilibrium assignment of the stable improvement cycles mechanism is Pareto inferior to the assignment from deferred acceptance when students are strategic.

[7] Balinski and Sönmez (1999) and Abdulkadiroğlu and Sönmez (2003) also interpet stability on equity grounds.

a centralized matching process.

1.8 Conclusion

Changes with high school assignment in New York provide a unique opportunity to study the impact of centralizing and coordinating school admissions on students using rich micro data on preferences, assignments, and enrollment. New York's new centralized mechanism led to a 40% increase in the rate at which students enroll in their assigned school. The increase in enrollment comes despite complaints that principals valued knowing where their school was ranked in student's preferences, and lose this ability in the new mechanism. The enrollment rates increase the most for students who seem likely to have more sophistication to navigate the old mechanism, weighing against the idea that multiple-offer systems are needed so that certain student groups stay in the public district.

The change in compliance rates with the new mechanism correlate well with the descriptive patterns we document using student preferences. Estimating school demand using data from the NYC mechanism raises some empirical challenges having to do with whether choices are truthful and families understand the process. Through a battery of alternative specifications, we find that most of our quantitative findings are robust to different ways to use information from student rankings. Student preferences place a heavy weight on distance, but students are far from ranking their closest school first and instead trade it off with other school attributes including size, demographic composition, and average test scores.

The main fact our analysis uncovers is that the new mechanism led families to travel 0.7 miles further to their assigned school, but the benefit of obtaining this assignment outweighs the cost of additional travel. This phenomenon may be driven by the fact that in the old mechanism, when students were left unassigned, they were administratively assigned to the closest school possible, which may have not been the best fit for students. Nearly all student demographic groups experience an increase in distance, but this increase is more than compensated by a more preferred school for these groups. Our preference estimates imply that the average student gains the equivalent of 0.25 miles in the new mechanism, with 0.96 miles directly coming from obtaining a more preferred school. Students from all boroughs, demographic groups, and baseline achievement categories obtain on average a more preferred assignment from the new mechanism. Even though some students will inevitably be assigned to less desirable schools, the heterogeneity in preferences generates a significant role for match-specific components of the assignment and highlights the importance of a good assignment mechanism for generating allocative efficiencies.

The findings reported here may be relevant in other resource allocation settings with uncoordinated mechanisms as in New York City's old system. As we have already discussed, authorities in London have adopted essentially the same reforms a few years after New York City did. It is of course possible that the impacts that we measure would be different if a school district's old mechanism did not have the same features as New York. For instance, in 2005, Boston switched assignment algorithms within centralized mechanisms, where incentives, not congestion, were a key motivation for the new mechanism. Another important aspect of New York City is that the new mechanism is based on student-proposing deferred acceptance algorithm where many schools give priority to applicants residing in the same borough as the school. It is possible that the effects we document differ for other centralized mechanisms with distinct matching algorithms or varied school priorities. For instance, in New Orleans in 2012, the Recovery School District went from a completely decentralized mechanism to one based on Gale's Top Trading Cycles mechanism, which may further encourage movement across the district.

Finally, it's worth emphasizing that our analysis has focused on the allocative aspects of school choice and different ways to assign students to schools. An interesting question is whether allocative changes contribute to changes in the productive dimensions of assignment. That is, do different student assignment protocols impact the downstream outcomes of students? This far more difficult question requires understanding the link between choices and the schools that produce downstream impacts, but we hope to examine it in future work.

Chapter 2

An Empirical Model of the Medical Match

2.1 Introduction

Each year, the placement of about 25,000 medical residents and fellows is determined via a centralized clearinghouse known as National Residency Matching Program (NRMP) or "the match." During the match, applicants and residency programs list their preferences over agents on the other side of the market, and a stable matching algorithm uses these reported ranks to assign applicants to positions. Agents on both sides of the market are heterogeneous but salaries paid by residency programs are not individually negotiated with residents. Therefore, preferences of residents and programs, rather than prices, determine equilibrium outcomes. The medical match is iconic for the stable matching literature, but with few exceptions this literature has been primarily theoretical. Particularly, there is little evidence on the effects of government policies or the design of the market. These interventions can substantially affect the physician workforce in the United States because medical residents are a key component of current and future physician labor.[1]

This paper develops a new techniques for recovering the preferences of both the resi-

[1] According to the "2011 State Physician Workforce Data Book" (ww.aamc.org/workforce), in 2010, 678,324 physicians were reported as actively involved in patient care, whereas 110,692 residents and fellows were in training programs.

dency programs and residents (market primitives) using data only on final matches. The method may be useful for studying other matching markets because data on matches is common compared to stated preferences. As in the medical match, these primitives are important determinants of outcomes in matching markets when agents are heterogeneous and prices are not highly personalized. Examples include schooling, colleges and many high-skilled labor markets.

I estimate the model using data from the market for family medicine residents in the U.S. to empirically analyze two issues that have received particular attention from academic researchers as well as policy makers. First, I investigate the antitrust allegation that the centralized market structure is responsible for the low salaries paid to residents. The plaintiffs in a 2002 lawsuit argued that the match limited the bargaining power of the residents because salaries are set before ranks are submitted. They reasoned that a ”traditional market” would allow residents to use multiple offers and wage bargaining to make programs bid for their labor. Using a perfect competition model as the alternative, they argued that the large salary gap between residents and nurse practitioners or physician assistants is a symptom of competitive restraints imposed by centralization. Although the lawsuit was dismissed due to a legislated congressional exception, it sparked an academic debate on whether inflexibility results in low salaries (Bulow and Levin, 2006; Kojima, 2007) . Observational studies of medical fellowship markets do not find an association between low salaries and the presence of a centralized match (Niederle and Roth, 2003a, 2009) . While these studies strongly suggest that the match is not the primary cause of low salaries in this market, they do not explain why salaries in decentralized markets remain lower than the perfect competition salary benchmark suggested by the plaintiffs. I use a stylized theoretical model to show that residents’ preferences for programs result in an ”implicit tuition” that depresses salaries in a decentralized market. I then quantify the magnitude of this markdown using estimates from the empirical model.

Second, I study policy interventions for lowering the perceived under-supply of residents and physicians in rural areas of the U.S. Although a fifth of the U.S. population lives in rural areas, less than a tenth of physicians practice in rural communities (Rosenblatt and Hart, 2000) . The Patient Protection and Affordable Care Act of 2010 addresses the shortage of rural physicians by funding an increase in the number of residency programs in rural areas, redistributing unused Medicare funds originally allocated for residency training in urban hospitals, and increasing the funding of loan forgiveness programs used to recruit physicians to shortage areas. Broadly speaking, the act uses a combination of supply interventions and financial incentives to address the disparity in access to care. Such regulations are not unique to the United States. Recently, Japan reduced capacities in urban residency programs to mitigate their rural resident shortage (Kamada and Kojima, 2010) . Similar regulations affecting prices and quantities are common in a variety of matching markets but their effects on assignments are not well understood.[2]

Analyzing the general equilibrium effects of government policy as well as predicting outcomes under alternative market structures using counterfactual simulations require estimates of the preferences of both sides of the market. Direct data on these market primitives is frequently not available. Although the rank order lists submitted by residents and programs are collected by the NRMP, they are highly confidential. Preference lists may not even be collected in other labor or matching markets. When only data on final matches are available, it is not immediately clear how to use these data to estimate preferences.

This paper develops methods for estimating preferences using only data on final matches. The techniques apply to a many-to-one two-sided matching market with low frictions. Motivated by properties of the mechanism used in the medical match, I assume that the final matches are pairwise stable (Roth and Sotomayor, 1992) . According to this equilibrium concept, no two agents on opposite sides of the market prefer each other over their match partners at pre-determined salary levels. Following the discrete choice literature, I model the preferences of each side of the market over the other as a function of characteristics of residents and programs, some of which are known to market participants but not to the econometrician. I use the pure characteristics model of Berry and Pakes (2007) for the preferences of residents for programs. This model allows for substantial heterogeneity in the preferences. However, a similarly flexible model for the program's preferences for residents raises identification issues and other methodological difficulties due to multiple equilibria. In the medical residency market, anecdotal evidence suggests that residents are

[2]Tuition regulations in public universities and financial aid programs are a salient example of price interventions in matching markets. Schooling reforms establishing new public schools or closing dysfunctional school programs are common interventions that directly affect supply.

largely vertically differentiated in skill because academic record and clinical performance are the main determinants of a resident's desirability to a program.[3] These factors are not observed in the dataset but should be accounted for. I therefore restrict attention to a model in which the programs' preferences for residents are homogenous and allow for an unobservable determinant of resident skill. The assumption also implies the existence of a unique pairwise stable match and a computationally tractable simulation algorithm.

The empirical strategy must confront the fact that ”choice sets” of agents in the market are not observed because they depend on the preferences of other agents in the market. Instead of a standard revealed preference approach, I identify the model using observed sorting patterns between resident and program characteristics, and information only available in an environment with many-to-one matching. For example, residents from more prestigious medical schools sort into larger hospitals if medical school prestige is positively associated with human capital and hospital size is preferable. If residents from prestigious medical schools have higher human capital, they will not sort into larger hospitals if small hospitals are preferable. Furthermore, the degree of assortativity between medical school prestige and hospital size increases with the weight agents place on these characteristics when making choices. However, sorting patterns alone are not sufficient for determining the parameters of the model. A high weight on medical school prestige and a low weight on hospital size results in a similar degree of sorting as a high weight on hospital size and low weight on medical school prestige. Fortunately, data from many-to-one matches has additional information that assists in identification. In a pairwise stable match, all residents at a given program must have similar human capital. Otherwise, the program can likely replace the least skilled resident with a better resident. Because the variation in human capital within a program is low, the variation in residents' medical school prestige within programs is small if medical school prestige is highly predictive of human capital. The within-program variation in medical school prestige decreases with the correlation of human capital with medical school prestige. Note that it is only possible to calculate the

[3] Conversations with Dr. Katz, Program Director of Internal Medicine Residency Program at Brigham and Women's Hospital, suggest that while programs have some heterogeneous preferences for resident attributes, the primarily trend is that better residents get their pick of programs ahead of less qualified residents. Further, academic and clinical record, and recommendation letters are the primary indicators used to determine resident quality.

within-program variation in a resident characteristic if many residents are matched to the same program. Finally, to learn about heterogeneity in preferences, I use observable characteristics of one side of the market that are excluded from the preferences of the other side. These exclusion restrictions shift the preferences of, say residents, without affecting the preferences of programs, thereby allowing sorting on excluded characteristics to be interpreted in terms of preferences.

I estimate the model using the method of simulated moments (McFadden, 1989; Pakes and Pollard, 1989) , and data from the market for family medicine residents between 2003 and 2010. Approximately 430 programs and 3,000 medical residents participate in this market each year. Moments used in estimation include summaries of the sorting patterns observed in the data and the within-program variation in observable characteristics of the residents. The small number of markets and the interdependence of observed matches creates additional challenges for estimation and inference. Instead of considering asymptotic approximations based on independently sampled matches or many markets, I mimic a data generating process in which the market grows in size. The characteristics of the market participants are drawn iid from a population distribution and the pairwise stable match for the realized market is observed. The dependence of matches on characteristics of all agents necessitates the use of a parametric bootstrap for constructing confidence sets for the estimated parameter.[4]

I show how to modify the model to correct for potential endogeneity between salaries and unobserved program characteristics. The technique is based on a control function approach and relies on the availability of an instrument that is excludable from the preferences of the residents (see Blundell and Powell, 2003; Heckman and Robb, 1985; Imbens and Newey, 2009). This approach can be used in other applications in labor markets where endogeneity may arise from compensating differentials or other influences on equilibrium wages. For this setting, I construct an instrument using Medicare's reimbursement rates to competitor residency programs, which are based on regulations enacted in 1985. The results from the instrumented version of the model are imprecise but indicate that salaries

[4]Agarwal and Diamond (2013) studies asymptotic theory for a single large market and the special case with homogeneous preferences on both sides. Monte Carlo evidence suggests that the root mean squared error drops with sample size and confidence sets have close to correct coverage.

are likely positively correlated with unobservable program quality.

I assess the fit of the model, both in-sample and out-of-sample. The out-of-sample fit uses the most recent match results, taken from the 2011-2012 wave of the census. These data were not accessed until estimates were obtained. The observed sorting patterns for resident groups mimic those predicted by the model, both in-sample and out-of sample, suggesting that the model is appropriately specified.

Counterfactual simulations are used to analyze the issues related to the lawsuit and policy interventions for rural training. In the lawsuit, the plaintiffs used a perfect competition model to argue that residents' salaries are lower than those paid to substitute health professionals because of the match. This reasoning does not account for the effects of the limited supply of heterogeneous programs and residents. A shortage of desirable residency programs due to accreditation requirements may lower salaries at high quality programs. Symmetrically, highly skilled residents can bargain for higher compensation because they are also in limited supply. Equilibrium salaries under competitive negotiations are influenced by both of these forces. I use a stylized model to show that when residents value program quality, salaries in every competitive equilibrium are well below the benchmark level suggested by the plaintiffs. The markdown is due to an implicit tuition arising from residents' willingness to pay for training at a program, and is in addition to any costs of training passed through to the residents. I estimate an average implicit tuition of at least \$23,000, with larger implicit tuitions at more desirable programs. Although imprecisely estimated, estimates from models using wage instruments are much higher, at \$43,000. The results weigh against the plaintiffs' claim that in the absence of competitive restraints imposed by the match, salaries paid to residents would be equal to the marginal product of their labor, close to salaries of physician assistants and nurse practitioners. At a median salary of \$86,000, physician assistants earn approximately \$40,000 more than medical residents. The upper-end of the estimated implicit tuition can explain this difference. These results imply that the low salaries observed in this market and those observed by Niederle and Roth (2003a, 2009) in the related medical fellowship markets without a match are due to the implicit tuition, not the design of the match.

Second, regulations aimed at increasing the number of residents in rural areas also affect sorting through general equilibrium effects. A reduction in urban training positions displaces residents who can in-turn displace other residents who get assigned elsewhere. Financial incentives for rural training and increases in the number of positions in rural areas cause similar re-sorting. The net impact of policy interventions is a function of the preferences of both residents and programs as well as the overall composition of the market. Using estimates from the model, I show that financial incentives have only a moderate effect on the number of residents matched to rural programs. An incentive of \$10,000 per year increases the number of residents in rural areas by about 17, or 5% of the total number of positions in rural programs. At a total cost of \$3.3 million, each additional resident in a rural program costs \$200,000 on average. This large per-resident cost arises because most of the incentives accrue to residents occupying positions that would have been filled without the incentive. Only 7.7% of rural residency positions are unfilled to begin with, which allows little scope for salary incentives to increase numbers. Instead, the primary impact of this policy is an increase in the quality of residents in rural areas. As expected, policy interventions directed at the supply of positions are more effective at increasing the number of residents placed at rural programs. Depending on the design of the regulation, supply interventions can either increase or decrease the quality of residents matched at rural programs through general equilibrium re-sorting effects. I find that a policy reducing positions offered in urban programs forces residents into rural programs, but due to re-sorting, does not significantly lower the quality of residents matched at rural programs. An increase in the number of positions offered in rural programs, on the other hand, increases the quality of residents training in rural communities through disproportional take-up in higher quality rural programs.

The empirical methods in this paper contribute to the recent literature on estimating preference models using data from observed matches and pairwise stability in decentralized markets.[5] The majority of papers focus on estimating a single aggregate surplus that is divided between match partners. Chiappori et al. (2011), Galichon and Salanie (2010), among others, build on the seminal work of Choo and Siow (2006) for studying transfer-

[5]See Fox (2009) for a survey. The approach of using pairwise stability in decentralized markets may yield a good approximation of market primitives if frictions are low. Many studies are devoted to understanding the role of search frictions as a determinant of outcomes in decentralized labor and matching markets (Mortensen and Pissarides, 1994; Postel-Vinay and Robin, 2002; Roth and Xing, 1994; Shimer and Smith, 2000).

able utility models of the marriage market in which an aggregate surplus is split between spouses. Fox (2008) proposes a different approach for estimation, also for the transferable utility case, with applications in Bajari and Fox (2005), among others. Sorensen (2007) is an example that estimates a single surplus function, but in a non-transferable utility model. Another set of papers measures benefits of mergers using similar cooperative solution concepts (Akkus et al., 2012; Gordon and Knight, 2009; Uetake and Watanabe, 2012; Weese, 2008) . A common data constraint faced in many of these applications is that monetary transfers between matched partners are often not observed, so the possibility of estimating two separate utility functions is limited.

Since salaries paid by residency programs are observed, this paper can estimate preferences of each of the two sides of the market, with salary as a (potentially endogenous) additional characteristic that is valued by residents. I use a non-transferable utility model because the salary paid by a residency program is pre-determined. Similar models are estimated by Logan et al. (2008) and Boyd et al. (2003) , although in decentralized markets, with the goal of measuring preferences for various characteristics. Logan et al. (2008) proposes a Bayesian method for estimating preferences for mates in a marriage market with no monetary transfers. Boyd et al. (2003) uses the method of simulated moments to estimate the preferences of teachers for schools and of schools for teachers. Both papers use only sorting patterns in the data to estimate and identify two sets of preference parameters. Agarwal and Diamond (2013) prove that even under a very restrictive model with no preference heterogeneity on either side of the market, sorting patterns alone cannot identify the preference parameters of the model. Such non-identification can yield unreliable predictions for both counterfactuals studied in this paper. To solve this problem, I leverage information made available through many-to-one matches, in addition to sorting patterns, for identifying two distributions of preferences.

The results on equilibrium salaries paid to residents may also be of independent interest for their analysis of labor markets with compensating differentials, especially those with on-the-job training. It is well known that compensating differentials can be an important determinant of salaries in labor markets (Rosen, 1987) . Stern (2004), for instance, finds that scientists often accept lower salaries from firms that allow their employees to publish research. Previous theoretical work on markets with on-the-job training has used perfect competition models to show that salaries are reduced by the marginal cost of training (Becker, 1975; Rosen, 1972). Counterfactuals in this paper using the competitive equilibrium model compute an implicit tuition at residency programs, which a markdown due to the value of training that is in addition to costs of training passed through to the resident.

The paper begins with a description of the market for family medicine residents and the sorting patterns observed in the data (Section 2.2). Sections 2.3 through 2.7 present the empirical framework used to analyze this market, the identification strategy, the method for correcting potential endogeneity in salaries, the estimation approach, and parameter estimates, respectively. These sections omit details relevant exclusively to the applications related to the lawsuit and the analysis of policy for encouraging rural training. Background for each issue is presented along with counterfactual simulations in Sections 2.8 and 2.9 respectively. All technical details are relegated to appendices.

2.2 Market Description and Data

This paper analyzes the family medicine residency market from the academic year 2003-2004 to 2010-2011. The data are from the National Graduate Medical Education Census (GME Census) which provides characteristics of residents linked with information about the program at which they are training.[6] Family medicine is the second largest specialty, after internal medicine, constituting about one eighth of all residents in the match. Graduates from family medicine residency programs provide the bulk of medical care in rural United States (Rosenblatt and Hart, 2000).

I focus on five major types of program characteristics: the prestige/quality of the program as measured by NIH funding of a program's major and minor medical school affiliates;[7] the size of the primary clinical hospital as measured by the number of beds; the

[6] I consider all non-military programs participating in the match, accredited by the Acceditation Council of Graduate Medical Education and not located in Puerto Rico. I restrict attention to residents matched with these programs. Detailed description of all data sources, construction of variables, sample restrictions and the process used to merge records are in Appendix B.5. Data on matches from the Graduate Medical Education Database, Copyright 2012, American Medical Association, Chicago, IL.

[7] Major affiliates of a program are directly affiliated medical schools of a program's primary clinical hospital. Other medical school affiliations between programs and medical schools, via secondary rotation sites

Medicare Case Mix Index as a measure of the diagnostic mix a resident is exposed to; characteristics of program location such as the median rent in the county a program is located in and the Medicare wage index as a measure of local health care labor costs; and the program type indicating the community and/or university setting and/or rural setting of a program.

Table 2.1 summarizes the characteristics of programs in the market. The market has approximately 430 programs, each offering approximately eight first-year positions. Except for program type (community/university based), there is little annual variation in the composition of programs in the market. Salaries paid to residents have roughly kept up with inflation with a distribution compressed around $47,000 in 2010 dollars.[8]

In general, rural programs are smaller than urban programs. They typically consist of about five residency positions, are at smaller hospitals as measured by the number of beds, and are affiliated with medical schools with lower NIH funding. Even though family medicine physicians provide the majority of care in rural communities where 20% of the US population resides, only about 10% of residency positions in this specialty are in rural settings.

For residents, the data contains information on their medical degree type,characteristics of graduating medical school and city of birth. Table 2.2 describes the characteristics of residents matching with family medicine programs. The composition of this side of the market has also been stable over this sample period with only minor annual changes. A little less than half the residents in family medicine are graduates of MD granting medical schools in the US. A large fraction, about 40%, of residents obtained medical degrees from non-US schools while the rest have US osteopathic (DO) degrees.[9] One in ten US born medical residents are born in rural counties.

---

or other affiliates of the primary clinical site, are categorized as minor. See data appendix for details.

[8]Resident salaries after the first year is highly correlated with the first year salary with a coefficient that is close to one and a R-squared of 0.8 or higher.

[9]As opposed to allopathic medicine, osteopathy emphasizes the structural functions of the body and its ability to heal itself more than allopathic medicine. Osteophathic physicians obtain a Doctor of Osteopathy (DO) degree and are licensed to practice medicine in the US just as physicians with a Doctor of Medicine (MD) degree.

Table 2.1: Program Characteristics

	2010-2011				2003-2004 to 2010-2011			
	All Programs N = 428		Rural Programs N = 63		All Programs N = 3,441		Rural Programs N = 481	
	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
First Year Salary (2010 dollars)	47,331	2,953	47,234	2,598	46,394	3,239	46,259	2,882
Program Size	7.70	2.83	5.30	2.64	7.57	2.77	5.25	2.44
Number of Matches	7.36	2.93	4.92	2.77	7.01	2.92	4.72	2.51
NIH Funding (Major Affil., bil $)	58.85	86.04	46.51	76.35	56.97	84.85	37.71	61.48
NIH Funding (Minor Affil., bil $)	57.98	77.21	40.98	75.02	53.25	83.87	37.34	82.22
Beds (Primary Inst)	421.54	284.15	240.47	151.81	418.41	273.17	257.54	150.29
Community Based Program	0.25	0.43	0.29	0.46	0.33	0.47	0.39	0.49
Community-University Program	0.62	0.49	0.63	0.49	0.54	0.50	0.53	0.50
University Based Program	0.13	0.34	0.08	0.27	0.12	0.33	0.07	0.26
Number of Interviews	63.38	31.10	37.62	22.07	55.56	30.17	31.91	20.14
Medicare Case Mix Index (Prim. Inst)	1.61	0.23	1.52	0.20	1.57	0.22	1.50	0.22
Medicare Wage Index (Prim. Inst)	1.00	0.14	0.93	0.11	1.01	0.14	0.93	0.10

Notes: Details on the construction of variables and the rule for classifying a program as rural is provided in the data appendix. Statistics on interviews and Medicare fields reported conditional on non-missing data. Less than 2% of the data on these fields is missing. NIH fund statistics are reported only for programs with NIH funded affiliates. About 35% of the programs have no NIH funded major affiliates, while about 46% have no minor affiliates. About 8% of programs have no NIH funded medical school affiliates. All other characteristics have full coverage.

Table 2.2: Resident Characteristics

	2010-2011 N = 3,148		2003-2004 to 2010-2011 N = 24,115	
	Mean	Std	Mean	Std
Allopathic/MD Graduate	0.45	0.50	0.45	0.50
Osteopathic/DO Graduate	0.15	0.36	0.14	0.34
Foreign Medical Graduate	0.39	0.49	0.41	0.49
NIH Funding (MD grads, mil $)	83.26	82.42	84.08	83.96
Median MCAT Score (MD grads)	31.24	2.25	31.31	2.20
US born Foreign Graduate	0.12	0.33	0.09	0.29
Rural Born Resident	0.11	0.31	0.10	0.30

Notes: Details on the construction of variables provided in the data appendix. A resident is classified as rural born if her city of birth is not in an MSA. City of birth data is unreliable for about 7.3% residents - rural born is coded as missing for these residents. Country of birth is not known for 14.6% of residents, and are treated as foreign graduates not born in the US.

2.2.1 The Match

A prospective medical resident begins her search for a position by gathering information about the academic curriculum and terms of employment at various programs from an online directory and official publications. Subsequently, she electronically submits applications to several residency programs which then select a subset of applicants to interview. On average, approximately eight residents are interviewed per position (Table 2.1). Anecdotal evidence suggest that during or after interviews, informal communication channels actively operate allowing agents on both sides of the market to gather more information about preferences. Finally, residency programs and applicants submit lists stating their preferences for their match partners. The algorithm described in Roth and Peranson (1999) uses these rank order lists to determine the final match. The terms of participating in the match create a commitment by both the applicant and the program to honor this assignment. Programs do not individually negotiate salaries with residents during this process.

The centralized market for medical residents was established in the 1950s to create a uniform transaction date, primarily as a remedy for discernible inefficiencies caused by early and exploding offers (Roth, 1984; Roth and Xing, 1994). In 1998, the clearinghouse was redesigned amid concerns that the existing design was not in the best interest of applicants and to lower difficulties with solving colocation problems for residency applicants married to other applicants (Roth and Peranson, 1999). The algorithm currently in use substantially reduces incentives for residents and programs to rematch by producing a match in which no applicant and program pair could have ranked each other higher than their assignments. It is adapted from the instability-chaining algorithm of Roth and Vande Vate (1990) and shares features with the applicant proposing deferred acceptance algorithm introduced by Gale and Shapley (1962).

A few positions are filled before the match begins and some positions not filled after the main match are offered in the ”scramble.” During the scramble, residents and programs are informed if they were not matched in the main process and can use a list of unmatched agents to contract with each other.[10]

[10] A new managed process called the Supplemental Offer Acceptance Program (SOAP) replaced the scramble in 2012. A total of 142 positions in family medicine (approximately 5%) were filled through this process.

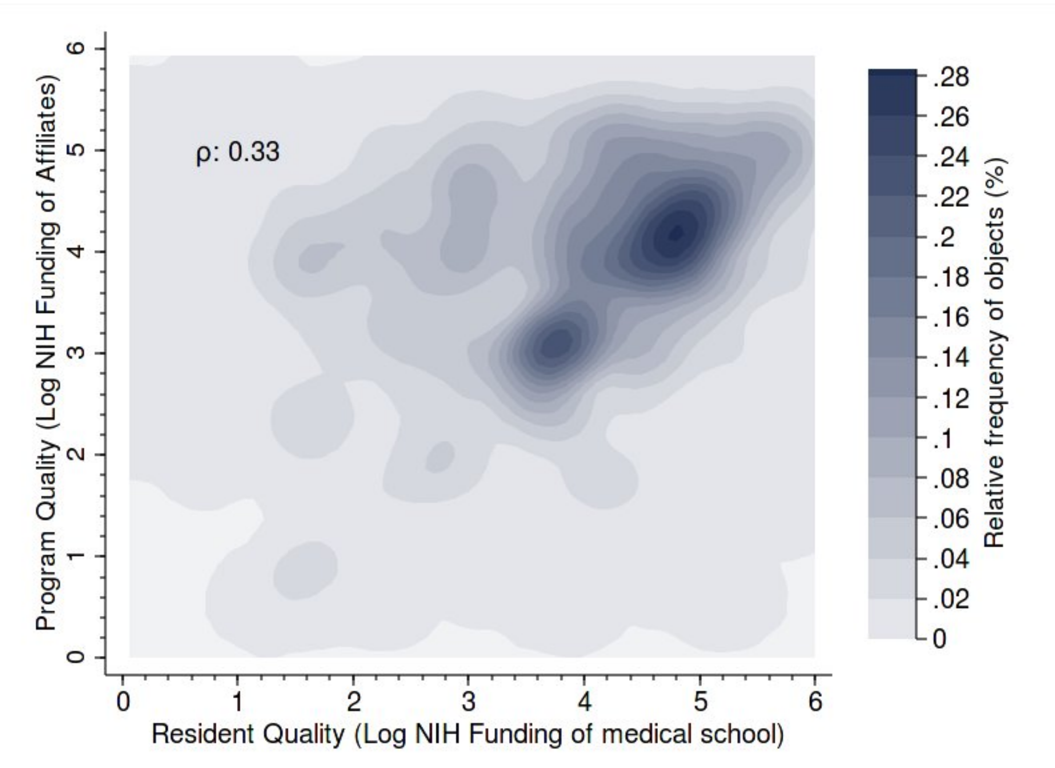
2.2.2 Descriptive Evidence on Sorting

Motivated by the properties of the match, the empirical strategy uses pairwise stability to infer parameters of the model by taking advantage of sorting patterns between resident and program characteristics observed in the data and features of the many-to-one matching structure to infer preferences. I defer the discussion of the many-to-one aspect to Section 2.4.2.

There is a significant degree of positive assortative matching between measures of a resident's medical school quality and that of a program's medical school affiliates. Figure 2.1 shows the joint distribution of NIH funding of a resident's medical school and of the affiliates of the program with which she matched. Residents from more prestigious medical schools, as measured by NIH funding, tend to match to programs with more prestigious medical school affiliates. Table 2.3 takes a closer look at this sorting using regressions of a resident's characteristic on the characteristics of programs with which she is matched. The estimates confirm the general trend observed in Figure 2.1. Programs that are associated with better NIH funded medical schools tend to match with residents from better medical schools as well, whether the quality of a resident's medical school is measured by NIH funding, MCAT scores of matriculants, or the resident having an MD degree rather than an osteopathic or foreign medical degree. This observation also holds true for programs at hospitals with a higher Medicare case mix index as well. Rent is positively associated with resident quality, potentially because cities with high rent may also be the ones that are more desirable to train or live in. Also note that the coefficient on the rural program dummy is not statistically significant. *Ceteris paribus*, rural programs are not matched with significantly lower quality residents than urban programs. Further, statistics from Table 2.1 show that about 90% of positions in rural programs are filled, while 93% are at urban programs. These findings are consistent with survey evidence in Rosenblatt et al. (2006), which shows that rural training programs are matched with residents of a similar type as urban programs.[11]

The scramble was likely of a similar size in the earlier years. See Signer (2012) (accessed June 12, 2012).

[11] Unlike Rosenblatt et al. (2006), my analysis includes positions in rural residency training track programs that are satellites of urban host programs.



Notes: Darker regions depict higher density. Density calculated using two-dimensional bandwidths using a quartic kernel and a bandwidth of 0.6. Log NIH Fund of Affiliates is the log of the average of NIH funds at major and minor affiliates. Sample restricted academic year 2010-2011 and programs with at least one NIH funded affiliate and residents from NIH funded medical schools.

Figure 2.1: Assortative Matching between Programs and Residents

Table 2.3: Sorting between Residents and Programs

	Log NIH Fund (MD)	Median MCAT (MD)	MD Degree	DO Degree
	(1)	(2)	(3)	(4)
Log NIH Fund (Major)	0.3724***	0.0154***	0.0462***	0.0025
	(0.0119)	(0.0007)	(0.0032)	(0.0022)
Log NIH Fund (Minor)	0.1498***	0.0084***	0.0208***	0.0048*
	(0.0137)	(0.0008)	(0.0040)	(0.0028)
Log # Beds	-0.0972***	-0.0021	-0.0104	-0.0098**
	(0.0221)	(0.0014)	(0.0064)	(0.0045)
Rural Program	-0.0687	-0.0040	-0.0010	0.0138*
	(0.0437)	(0.0027)	(0.0117)	(0.0082)
Log Case-Mix Index	0.1894**	0.0136**	0.4670***	0.0574***
	(0.0940)	(0.0058)	(0.0255)	(0.0179)
Log First-Year Salary	0.0126	0.0590***	0.3001***	0.0969***
	(0.1717)	(0.0106)	(0.0467)	(0.0327)
Log Rent	0.4612***	0.0727***	0.1811***	-0.0012
	(0.0600)	(0.0037)	(0.0168)	(0.0118)
Observations	10,842	10,872	23,984	23,984
R-squared	0.1318	0.1282	0.0381	0.0079

Notes: Linear regression of resident's graduating school characteristic on matched program characteristics. Samples pooled from the academic years 2003-2004 to 2010-2011. Column (1) restricts to the set of residents graduating from medical schools with non-zero average annual NIH funding. Column (2) restricts to the subset of residents with MD degrees from institutions reporting a median MCAT score in the Medical School Admission Requirements in 2010-2011. Columns (3) and (4) include all residents. See data appendix for description of variables. All specifications include dummy variables for programs with no NIH funding at major affiliates, no NIH funding at minor affiliates and a missing Medicare ID for the primary institution. Standard errors in parenthesis. Significance at 90% (*), 95% (**) and 99% (***) confidence.

Table 2.4: Geographical Sorting between Residents and Programs

	Log NIH Fund (Major) (1)	Log NIH Fund (Minor) (2)	Log # Beds (3)	Log Case Mix Index (4)	Rural Program (5)
Log NIH Fund (MD)	0.4058***	0.1555***	-0.0213***	-0.0002	-0.0110***
	(0.0124)	(0.0116)	(0.0046)	(0.0011)	(0.0023)
Log Median MCAT (MD)	0.6953***	0.4704***	0.0830**	0.0023	-0.0877***
	(0.1009)	(0.0914)	(0.0364)	(0.0091)	(0.0184)
US Born (For)	-0.0711*	-0.1032***	-0.0025	0.0186***	0.0141*
	(0.0374)	(0.0366)	(0.0143)	(0.0036)	(0.0072)
Match in Med Sch. State	-0.4463***	-0.2646***	0.0468***	-0.0057*	0.0111*
	(0.0322)	(0.0303)	(0.0121)	(0.0030)	(0.0061)
Match in Birth State	-0.0038	0.0197	-0.0376***	-0.0075***	-0.0115**
	(0.0285)	(0.0264)	(0.0105)	(0.0026)	(0.0053)
Rural Born Resident					0.0714***
					(0.0066)
Observations	15,394	13,099	24,115	23,652	24,115
R-squared	0.1211	0.0299	0.0052	0.0167	0.0101

Notes: Linear regression of characteristics of program or program affiliates on characteristics of matched residents. Samples pooled from the academic years 2003-2004 to 2010-2011. Column (1) restricts the sample to the set of programs with major affiliates that have positive NIH funding. Column (2) restricts the sample to the set of programs with a minor affiliate with non-zero NIH funding. Column (3) and column (5) includes all programs. Columns (4) excludes programs for which the Medicare ID is missing. All specifications have medical school type dummies and a dummy for residents graduating from MD medical schools without NIH funding. Column (5) includes a dummy for non-reliable city of birth information for US born residents. See data appendix for description of variables. Standard errors in parenthesis. Significance at 90% (*), 95% (**) and 99% (***) confidence.

To highlight the geographical sorting observed in the data, Table 2.4 regresses characteristics of a resident's matched program on her own characteristics and indicators of whether the program is in her state of birth or medical school state. Residents that match with programs in the same state as their medical school tend to match with less prestigious programs, as measured by the NIH funds of a program's affiliates. Residents also match with programs that are at larger hospitals and have lower case mix indices. Column (5) shows that rural-born residents are about seven percentage points more likely to place at rural programs than their urban-born counterparts.

Since these patterns arise from the mutual choices of residents and programs, estimates from these regressions are not readily interpretable in terms of the preferences of either side of the market. In particular, none of the coefficient estimates in these regressions can be interpreted as weights on characteristics in a preference model. The next section develops a model of the market that is estimated using these patterns in the data.

2.3 A Framework for Analyzing Matching Markets

This section presents the empirical framework for the model, treating salaries as exogenous. I demonstrate how an instrument can be used to correct for correlation between salaries and unobserved program characteristics in Section 2.5.

2.3.1 Pairwise Stability

I assume that the observed matches are pairwise stable with respect to the true preferences of the agents, represented with $\succeq_k$ for a program or resident indexed by k. Each market, indexed by t, is composed of N_t residents, $i \in \mathcal{N}_t$ and J_t programs, $j \in \mathcal{J}_t$. The data consists of the number positions offered by program j in each period, denoted c_{jt}, and a match, given by the function $\mu_t : \mathcal{N}_t \to \mathcal{J}_t$. Let $\mu_t^{-1}(j)$ denote the set of residents program j is matched with.

A pairwise stable match satisfies two properties for all agents i and j participating in market t:

1. Individual Rationality

- For residents: $\mu_t(i) \succeq_i \phi$ where ϕ denotes being unmatched.
- For programs: $|\mu^{-1}(j)| \leq c_{jt}$ and $\mu_t^{-1}(j) \succeq_j \mu_t^{-1}(j) \setminus \{i\}$ for all $i \in \mu_t^{-1}(j)$.

2. No Blocking: if $j \succ_i \mu_t(i)$ then

- If $|\mu_t(j)| = c_{jt}$, then for all $i' \in \mu_t^{-1}(j)$, $\mu_t(j) \succeq_j (\mu_t(j) \setminus \{i'\}) \cup \{i\}$
- If $|\mu(j)| < c_j$, then $\mu_t(j) \succeq_j \mu_t(j) \cup \{i\}$.

A pairwise stable need not exist in general or there may be multiple pairwise stable matches. The preference model described in the subsequent sections guarantees the existence and uniqueness of a pairwise stable match.

Individual rationality, also known as acceptability, implies that no program or resident would prefer to unilaterally break a match contract. Because I do not observe data on unmatched residents, I assume that all residents are acceptable to all programs and that all programs are acceptable to all residents. Almost all US graduates applying to family medicine residencies as their primary choice are successful in matching to a family medicine program, and the number of unfilled positions in residency programs in this speciality is under 10%.[12] The primary limitation this assumption is the inability to account for substitution into other professions or entry by new residents.

Under the no blocking condition, no resident prefers a program (to her current match) that would prefer hiring that resident in place of a current match if the program has exhausted its capacity. If the program a resident prefers is empty, the program would not like to fill the position with that resident.

Theoretical properties of the mechanism used by the NRMP guarantees that the final match is pairwise stable with respect to submitted rank order lists, but not necessarily with respect to true preferences. Strategic ranking and interviewing, especially in the presence of incomplete information, is likely the primary threat to using pairwise stability in this

[12]While residents may apply to many specialties in principle, data from the NRMP suggests that a typical applicant applies to only one or two specialties (except those looking for preliminary positions). A second specialty is often a "backup." Greater than 95% of MD graduates interested in family medicine, however, only apply to family medicine programs. Upwards of 97% residents that list a family medicine program as their first choice match to a family medicine program in the main match (See "Charting Outcomes in the Match" 2006, 2007, 2009, 2011, accessed June 12, 2012).

market.[13] The large number of interviews per position suggests that this may not be of concern in this market, however, it may be implausible in some decentralized markets.

This equilibrium concept also implicitly assumes that agents' preferences over matches is determined only by their match, not by the match of other agents. This restriction rules out the explicit consideration of couples that participate in the match by listing joint preferences.[14] According to data reports from the NRMP, in recent years only about 1,600 out of 30,000 individuals participated in the main residency match as part of a couple. I model all agents as single agents because data from the GME census does not identify an individual as part of a couple.

2.3.2 Preferences of the Residents

Following the discrete choice literature, I model the latent indirect utility representing residents' preferences $\succeq_i$ as a function $U(z_{jt}, \eta_{jt}, w_{jt}, \beta_i; \theta)$ of observed program traits z_{jt}, the program's salary offer w_{jt}, unobserved traits ξ_{jt}, and taste parameters β_i. I use the pure characteristics demand model of Berry and Pakes (2007) for this indirect utility:

$$u_{ijt} = z_{jt}\beta_i^z + w_{jt}\beta_i^w + \xi_{jt}. \tag{2.1}$$

In models that do not use a wage instrument, I assume that the unobserved traits ξ_{jt} have a standard normal distribution that is independent of the other variables. I normalize the mean utility to zero for $(z,\ w) = 0$. The scale and location normalizations are without loss in generality. The independence of ξ_{jt} from w_{jt} is relaxed in the model correcting for potential endogeneity in salaries.

Depending on the flexibility desired, β_i can be modelled as a constant, a function of observable characteristics x_i of a resident and/or of unobserved taste determinants η_i:

[13] The data and the approach does not make a distinction for positions offered outside the match or during the scramble. The no blocking condition should be a reasonable approximation for the positions filled before the match as it is not incentive compatible for the agents to agree to such arrangements if either side expects a better outcome after the match. The condition is harder to justify for small number of the positions filled during the scramble. Note, however, that residents (programs) that participate in the scramble should not form blocking pairs with the set of programs (residents) that they ranked in the main round.

[14] Couples can pose a threat to the existence of stable matches (Roth, 1984) although results in Kojima et al. (2010) suggest that stable matches exist in large markets if the fraction of couples is small.

$$\beta_i = x_i \Pi + \eta_i. \tag{2.2}$$

The taste parameters η_i are drawn from a mean-zero normal distribution with a variance that is estimated. The richest specification used in this paper allows for heterogeneity via normally distributed random coefficients for NIH funding at major affiliates, beds, and Case Mix Index. This specification also allows for preference heterogeneity for rural programs based on a rural or urban birth location of the resident and heterogeneity in preference for programs in the resident's birth state or medical school state through interaction of x_i and z_{jt}. These terms are included to account for the geographic sorting observed in the market.

The pure characteristics model implies that residents have tastes for a finite set of program attributes. It omits a commonly used additive ϵ_{ijt} term that is iid across residents, programs and markets. These discrete choice models implicitly assume tastes for programs through a characteristic space that increases in dimension with the number of programs. (Berry and Pakes, 2007) discuss some counter-intuitive implications of including an ϵ_{ijt} term on substitution patterns and welfare effects of changes in the number of programs.

2.3.3 Preferences of the Programs

Since the value produced by a team of residents at a program is not observed, I model residency program preferences through a latent variable. A very rich specification creates two extreme problems. On the one hand, a pairwise stable match need not exist if a program's preference for a given resident depends crucially on the other residents it hires. On the other hand, the number of stable matches can be exponentially large in the number of agents when programs have heterogenous preferences.[15] These problems are notwithstanding any difficulties one might face in identifying such a rich specification.

My conversations with residency program and medical school administrators suggests that programs broadly agree on what makes a resident desirable, and refer to a "peck-

[15]See Roth and Sotomayor (1992) for conditions of existence of a stable match in the college admissions problem. The multiplicity of the match implied by heterogeneous preference may not be particularly important from an empirical perspective. In simulations conducted with data reported to the NRMP, Roth and Peranson (1999) find that almost all of the residents are matched to the same program across all the stable matches.

ing order" for residency slots in which the best residents get their preferred choices over others. Anecdotal evidence also suggests that test scores in medical exams, clinical performance, and the strength of recommendation letters are likely the most important signals of a program's preference for a resident, but are not observed in the dataset (see Footnote 3). Therefore, I model a program's preference for a resident using a single human capital index $H(x_i, \varepsilon_i)$ that is a function of observable characteristics x_i of a resident and an unobservable determinant ε_i.[16] I use the parametric form

$$h_i = x_i\alpha + \varepsilon_i, \tag{2.3}$$

where ε_i is normally distributed with a variance that depends on the type of medical school a resident graduated from. For graduates of allopathic (MD) medical schools, x_i includes the log NIH funding and median MCAT scores of the resident's medical school. Characteristics also include the medical school type for residents, i.e. whether a resident earned an osteopathic degree (DO) or graduated from a foreign medical school. I also include an indicator for whether a resident that graduated from a foreign medical school was born in the US. Without loss of generality, the variance of ε_i for residents with MD degrees is normalized to 1 and the mean of h at $x = 0$ is normalized to zero.

This specification guarantees the existence and uniqueness of a stable match and a computationally tractable simulation algorithm that is described in Section 2.6.3.[17] Finally, Section 2.4.3 notes that identifying a model with heterogeneity relies on exclusion restrictions, in this case an observable program characteristic that is excluded from the preferences of the residents for programs.

Since heterogeneity in the preferences over residents is probable, bias in estimates may affect conclusions from counterfactual simulations. In particular, the analysis of interventions in rural residency training programs may be inaccurate if rural programs strongly

[16] The model only allows for ordinal comparisons between residents and is consistent with any latent output function $F_j\left(h_{i_1}, \ldots, h_{i_{c_j}}\right)$ from a team of residents $\left(i_1, \ldots, i_{c_j}\right)$ at program j that is strictly increasing in each of its components. An implicit restriction is that the preference for a resident does not depend on the other residents hired. The restriction may not be strong in this context becase programs cannot submit ranks that depend on the rest of the team.

[17] Existence follows since these preferences are *responsive*. The condition is similar to a substitutability condition. See Roth and Sotomayor (1992) for details. Uniqueness is a consequence of preference alignment. See Clark (2006) and Niederle and Yariv (2009).

prefer hiring rural-born residents. Appendix B.4.1 presents regressions showing that rural-born residents in rural programs are of similar (observable) quality as urban-born residents also matched to their residency programs. This suggests low heterogeneity in the preferences of programs, at least on this dimension.

2.4 Identification

In this section, I describe how the data provide information about preference parameters using pairwise stability as an assumption on the observed matches. The discussion also guides the choice of moments used in estimation. Standard revealed preference arguments do not apply because "choice-sets" of individuals are unobserved and determined in equilibrium.

Agarwal and Diamond (2013) study non-parametric identification in a single large market for a model without heterogenous preferences for programs. They find that having data from many-to-one matches rather than one-to-one matches is important from an empirical perspective. I intuitively describe the reason for this difference. A formal treatment of identification is beyond the scope of this paper.

The market index t is omitted in this section because all identification arguments are based on observing one market with many (interdependent) matches. For simplicity, I also assume that the number of residents is equal to the number of residency positions and treat all characteristics as exogenous. Identification of the case with endogenous salaries is discussed in Section 2.5, and does not require a reconsideration of arguments presented here.

2.4.1 Using Sorting Patterns: The Double-Vertical Model

Consider the simplified "double-vertical" model in which all residents agree upon the relative ranking of programs. In a linear parametric form for indirect utilities, preferences are represented with

$$\begin{aligned} u_j &= z_j\beta + \xi_j \\ h_i &= x_i\alpha + \varepsilon_i, \end{aligned}$$

where x_i and z_j are observed and ξ_j and ε_i are standard normal random variables, distributed independently of the observed traits. Assume the location normalizations $E[u_j|z_j = 0] = 0$ and $E[h_i|x_i = 0] = 0$.

A pairwise stable match in this model exhibits perfect assortative matching between u and h. Because the set of residents with a higher value of $x\alpha$ have a higher distribution of human capital, they are matched with more desirable programs. Conversely, programs with larger $z\beta$ are more likely to match with residents with higher human capital. The data exhibits positive assortativity between $x\alpha$ and $z\beta$. I now describe what learned from this sorting.

I begin with an example to show that a sign restriction on one parameter of the model is needed to interpret sorting patterns in terms of preferences. Consider a model in which x is a scalar measuring the prestige of a resident's medical school and z measures the size of the hospital with which a program is associated. In this example, residents from prestigious medical schools sort into larger hospitals if the human capital distribution of residents from more prestigious medical schools is higher and hospital size is preferable. However, this sorting may also have been produced by parameters under which residents from prestigious medical schools are less likely to have high human capital and smaller hospitals are preferable. This observation necessitates restricting one characteristic of either residents or programs to be desirable. Throughout the empirical exercises in this paper, I assume that residents graduating from more prestigious medical schools, as measured by the NIH funding of the medical school, are more likely to have a higher human capital index.[18] Under this sign restriction, the sorting patterns observed in Figure 2.1 can only be rationalized if a program's desirability is positively related to the NIH funding of its affiliates.

The sorting patterns can also allow us to determine whether $x\alpha = x'\alpha$ for $x \neq x'$ or conversely, if $z\beta = z'\beta$. Because $z\beta = z'\beta$, programs with characteristics z and z' are equally desirable to residents. Given a choice between these two programs, the unobservable characteristic ξ is used to break ties. For this reason, the distribution human capital of residents matched to the set of programs with observables z and z' are identical. Consider

[18]The sign restiction does not imply that all medical students at more prestigous medical schools have higher human capital index.

two types of programs, one at larger but less prestigious hospitals than another program at a smaller hospital. If residents trade-off hospital size for prestige, then the residents matched with these two hospital types have similar observable characteristics. Conversely, the distribution of observable quality of residents is higher at hospitals with characteristics z than at z' if $z\beta > z'\beta$. The nature of assortativity observed in the data thus informs us whether two observable types of residents or programs are equally desirable or not.

Agarwal and Diamond (2013) consider a more general model in which u and h are non-parametric functions of x and z respectively with additively separable errors ε and ξ. They prove that sorting patterns can be used to determine if x and x' are equally desirable.

2.4.2 Importance of Data from Many-to-One Matches

The preceding arguments using only sorting patterns do not contain information on the relative importance of observables on the two sides of the market. For intuition, consider an example in which x is a binary indicator that is equal to 1 for a resident graduating from a prestigious medical school and z is a binary indicator for a program at a large hospital. Assume that half the residents are from prestigious schools and half the programs are at large hospitals, and that medical school prestige and hospital size is preferred ($\alpha > 0$ and $\beta > 0$). Sorting patterns from such a model can be summarized in a contingency table in which residents from prestigious medical schools are systematically more likely to match with programs at large hospitals. For instance, consider the following table:

	$z = 1$	$z = 0$
$x = 1$	30%	20%
$x = 0$	20%	30%

These matches could result from parameters under which programs have a strong preference for residents from prestigious medical schools (large α) and residents have a moderate preference for large hospitals (small β). In this case, residents from more prestigious medical schools get their pick of programs, but often choose ones at small hospitals. On the other hand, the contingency table could have been a result of a strong preference for large hospitals (large β) but only a moderate preference for residents from prestigious medical schools (small α). There are a variety of intermediate cases that are indistinguishable from each other and either extreme. This ambiguity contrasts with discrete choice models using stated preference lists where the relationship between ranks and hospital size determines the weight on hospital size. Here, the degree of sorting between x and z cannot determine the weights on both characteristics because preferences of both sides determine final matches.

In addition to sorting patterns, data on many-to-one matches also determines the extent to which residents with similar characteristics are matched to the same program. In a pairwise stable match, two residents at the same program must have similar human capital irrespective of the program's quality. Otherwise, either the program could replace the lower quality resident with a better resident, or the higher quality resident is could find a more desirable program. Residents training at the same program have similar observables if x is highly predictive of human capital. Conversely, programs are not likely to match with multiple residents with similar observables if they placed a low weight on x. The variation in resident observable characteristics within programs is therefore a signal of the information observables contain about the underlying human capital quality of residents.[19]

This information is not available in a one-to-one matching market because sorting patterns are the only feature known from the data. Agarwal and Diamond (2013) formally shows that having data from many-to-one matches is critical for identifying the parameters of the model, and provides simulation evidence to illustrate the limitations of sorting patterns and the usefulness of many-to-one matching data.

Descriptive Statistics from Many-to-One Matching

Table 2.5 shows the fraction of variation in resident characteristics that is within a program. Notice that almost none of the variation in the gender of the resident is across programs. This fact suggests that gender does not determine the human capital of a resident. If gender were a strong determinant of a resident's desirability to a program, in a double-vertical model one would expect that programs would be systematically male or female

[19] An analogy with measurement error models to explans why many-to-one matches allow us to identify features we cannot in one-to-one match data. Since we expect that two residents matched to the same program are very similarly qualified, the observable quality of two doctors at the same program act like noisy measures of their identical true quality.

Table 2.5: Within Program Variation in Resident Characteristics

	Fraction of Variation Within Program-Year
Log NIH Fund (MD)	77.83%
Median MCAT (MD)	72.09%
US Born Foreign Graudate	79.01%
Osteopathic/DO Degree	85.16%
Foreign Degree	57.16%
Allopathic/MD Degree	64.81%
Female	96.40%

Notes: Each row reports $1 - R^2_{adj}$ from a separate linear regression of resident's graduating school characteristic absorbing the program-year fixed effects. Samples from the academic years 2003-2004 to 2010-2011. Samples for regressions with LHS variables Log NIH funding (MD), Median MCAT (MD) are restricted to the set of residents with non-missing values for the respective characteristic. Regression of US Born (For) restrict to graduates of foreign medical schools. Osteopathic/DO Degree, Foreign Degree, Allopathic/MD Degree are linear probability models estimated on the full sample.

dominated. Summaries of the other characteristics indicate that residents are more systematically sorted into programs where other residents have more similar qualifications. For instance, about 30% of the variation in the median MCAT score of the residents' graduating medical schools decomposes into across program variation. This statistic is higher for the characteristics foreign medical degree and MD degree.

Table 2.6 presents another summary from many-to-one matching based on regressing the leave one out mean characteristic of a resident's peer group in a program on the characteristics of the resident. Let $\bar{x}^{\mu}_{-i,1}$ be the average observable x_1 of resident i's peers for a match μ, i.e. $\bar{x}^{\mu}_{-i,1} = \frac{1}{|\mu^{-1}(\mu(i))|-1} \sum_{i' \in \mu^{-1}(\mu(i))} x_{i',1}$. I estimate the equation

$$\bar{x}_{-i,\mu} = x_i \lambda + e_i,$$

where x_i is resident i's observables. Not surprisingly, each regression suggests that a resident's characteristic is positively associated with the mean of the same characteristic of her peers. Viewing NIH funding, MCAT scores, and MD degree as quality indicators, there is a positive association between a resident's quality and the average quality of her peer group. Further, the moderately high R-squared statistics for these regressions suggest that resident characteristics are more predictive of her peer groups than what Table 2.5 might have suggested.

2.4.3 Heterogeneity in Preferences

I now discuss exclusion restrictions that can be used to learn about heterogeneity in preferences. Preferences based on observable characteristics of residents that do not affect their human capital index are reflected in heterogeneous sorting patterns for similarly qualified residents. Assume, for instance, that the birth location of a resident does not affect the preferences of programs for the resident. Under this restriction, the propensity of residents for matching to programs closer to their birthplace can only be a result of resident preferences, not the preferences of programs. Further, residents matching closer to home will do so at disproportionately lower quality programs since they trade off program quality with preferences for location.

The principle is similar to the use of variation excluded from one part of a system to identify a simultaneous equation model. The exclusion restriction in the example above

Table 2.6: Peer Sorting

	Peer Log NIH Fund	Peer Log MCAT	Peer Foreign Degree	Peer DO Degree	Peer MD Degree
	(1)	(2)	(3)	(4)	(5)
Log NIH Fund (MD)	0.2919***	0.0103***	-0.0249***	-0.0043**	0.0293***
	(0.0132)	(0.0026)	(0.0030)	(0.0019)	(0.0033)
Log Median MCAT (MD)	0.6449***	0.0874	-0.2000***	0.0165	0.1850***
	(0.1832)	(0.0750)	(0.0458)	(0.0247)	(0.0499)
US Born (For)	0.0403	0.0141	-0.1063***	0.0394***	0.0669***
	(0.0421)	(0.0103)	(0.0091)	(0.0050)	(0.0079)
Observations	19,830	19,845	24,066	24,066	24,066
R-squared	0.1280	0.6437	0.3632	0.0914	0.3197

Notes: Linear regression of average characteristics of peers on the characteristics of a resident. A peer of a resident is another resident matched to the same program as that resident in the academic cohort of said resident. The calculation of peer averages for a resident excludes the resident herself. Samples pooled from the academic years 2003-2004 to 2010-2011. Column (1) restricts the sample to the set of residents with at least one peer that graduated from a medical school with non-zero NIH funding. Column (2) restricts the sample to the set of residents with at least one peer that graduated from a medical school with non-missing MCAT Score. Peer averages for columns (1) and (2) are constructed only from peers with non-missing observations of these characteristics. Columns (3-5) considers all residents with at least one peer. All specifications have medical school type dummies and a dummy for residents graduating from MD medical schools without NIH funding. See data appendix for description of variables. Standard errors clustered at the program-year level in parenthesis. Significance at 90% (*), 95% (**) and 99% (***) confidence.

isolates a factor influencing the demand for residency positions without affecting the distribution of choice sets faced by residents. Conversely, one may use factors that influence the human capital index of a resident but not their preferences to obtain variation in choice sets of residents that is independent of resident preferences. Conlon and Mortimer (2010) use a similar source of variation arising from product availability to identify demand models with unobserved heterogeneity.

While only one restriction may suffice in theory, the empirical specifications in this paper use both restrictions. Ideally, one would be able to estimate preferences for programs that are heterogeneous across residents with different medical schools or skill levels. Richer specifications that allows for this type of preference heterogeneity are difficult to estimate because quality indicators of residents only include the medical school, and do not vary at the individual level. Even with more detailed information on residents, estimating the preferences for residents with low qualifications is likely to rely on parametric extrapolations from more qualified residents because of the limited set of choices faced by less skilled residents.

2.5 Salary Endogeneity

The salary offered by a residency program may be correlated with unobserved program covariates. For instance, programs with desirable unobserved traits may be able to pay lower salaries due to compensating differentials. Alternatively, desirable programs may be more productive or better funded, resulting in salaries that are positively associated with unobserved quality. One approach to correct for wage endogeneity is to formally model wage setting. I avoid this for several reasons. First, the allegation of collusive wage setting in the lawsuit is unresolved. Second, hospitals tend to set identical wages for residents in all specialties, suggesting that a full model should consider the joint salary setting decision across all residency programs at a hospital. Finally, a full model would need to account for accreditation requirements that require salaries to be ”adequate” for a resident’s living and educational expenses.[20]

2.5.1 A Control Function Approach

I propose a control function correction for bias due to correlation between salaries w_{jt} and program unobservables ξ_{jt} (see Blundell and Powell, 2003; Heckman and Robb, 1985; Imbens and Newey, 2009). The principle of the method is similar to that of an instrumental variables solution to endogeneity. It also relies on an instrument r_{jt} that is excludable from the utility function $U(\cdot)$. The instrument I use is described in the next section.

Consider the following linear function for the salary w_{jt} offered by program j in period t :

$$w_{jt} = z_{jt}\gamma + r_{jt}\tau + \nu_{jt}, \tag{2.4}$$

where z_{jt} are program observable characteristics, r_{jt} is the instrument, and ν_{jt} is an unobservable. Endogeneity of w_{jt} is captured through correlation between the unobservables ν_{jt} and ξ_{jt}. Equation (2.4) is analogous to the first stage of a two-stage least squares estimator and the equilibrium model of matches is analogous to the second stage.

The control function approach requires (ξ_{jt}, ν_{jt}) to be independent of (z_{jt}, r_{jt}). This assumption replaces weaker conditional moment restriction needed in instrumental variables approach.[21] Under this independence, although w_{jt} is not (unconditionally) independent of ξ_{jt}, it is conditionally independent of ξ_{jt} given ν_{jt} and z_{jt}. The control function approach uses a consistent estimate of ν_{jt} from the first stage as a conditioning variable in place of its true value.

Since ν_{jt} can be consistently estimated from equation (2.4) using OLS, treat it as any

[20]The ACGME sponsoring institution requirements state that ”Sponsoring and participating sites must provide all residents with appropriate financial support and benefits to ensure that they are able to fulfill the responsibilities of their educational programs.”

[21]Imbens (2007) discusses these independence assumptions at some length, noting that they are commonly made in the control function literature and are often necessary when dealing with a non-additive second stage. In this context, even though ξ_{jt} is additively separable from w_{jt}, the observed matches are not an additive function of ξ_{jt} and w_{jt}. This fact prohits the approach used in demand models pioneered by Berry (1994) and Berry et al. (1995), where an inversion can be used to to estimate a variable with a separable form in the unobserved characteristic and the endogenous variable.

other observed characteristic. As noted earlier, we need to allow for correlation between ν_{jt} and ξ_{jt} to build endogeneity of w_{jt} into the system. For tractability given the limited salary variation, I model the distribution of ξ_{jt} conditional on ν_{jt} as

$$\xi_{jt} = \kappa\nu_{jt} + \sigma\zeta_{jt}, \tag{2.5}$$

where $\zeta_{jt} \sim N(0,1)$ is drawn independently of ν_{jt} and (κ, σ) are unknown parameters. Substitute equation (2.5) to re-write equation (2.1) as

$$u_{ijt} = z_{jt}\beta_i^z + w_{jt}\beta_i^w + \kappa\nu_{jt} + \sigma\zeta_{jt}. \tag{2.6}$$

Since variation in w_{jt} given ν_{jt} and z_{jt} is due to r_{jt}, the assumptions above imply that ζ_{jt} is independent of w_{jt}, solving the endogeneity problem.

As a scale normalization, I set $\sigma = 1$. The term ζ_{jt} can arise from specification error and/or from unobservable determinants of salaries that do not directly affect the preferences of residents for a program. Note that the unobservable characteristic of the program ξ_{jt}, may be correlated across time through ν_{jt}. For instance, ν_{jt} may be the sum of a random effect ν_j^r that is constant over time for a given j and a per-period deviation ν_{jt}^d as long as each of the components is independent of (z_{jt}, r_{jt}).

While this linear specification may be difficult to justify from economic primitives, it may substantially reduce bias in estimates. Even in models of oligopolistic competition in which the price has a nonlinear relationship with unobservables and the characteristics of competing products, Yang et al. (2003) and Petrin and Train (2010) find that linear control functions can lead to significant reduction in bias. The restriction that w_{jt} does not depend on characteristics of other programs may not be particularly strong in this context. However, the single dimensional additive source of error, ν_{jt}, remains a strong assumption since it rules out heterogeneous effects of the instrument. It may be feasible to relax some parametric assumptions in equations (2.5) and (2.6) in settings with greater variation in the endogenous variable.

2.5.2 Instrument

Table 2.8 presents regression estimates of equation (2.5), except using a log-log specification so that coefficients can be interpreted as elasticities. The first four columns do not include the instrument r_{jt}, which is defined below. Columns (1) and (2) show limited correlation between salaries and observed program characteristics except rents and the Medicare wage index. The elasticity with respect to these two variables is small, at less than 0.15 in magnitude. This suggests that models that do not instrument for salaries may provide reasonable approximations for residents' preferences. To address potential correlation, I will also present estimates from specifications that use reimbursement rates for residency training at competitor hospitals as a wage instrument.

Medicare reimburses residency programs for direct costs of training based on cost reports submitted in the 1980s. Before the prospective payment system was established, the total payment made to a hospital did not depend on the precise classification of costs as training or patient care costs. The reimbursement system for residency training was severed from payments for patient care in 1985 because the two types of costs were considered distinct by the government. While patient care was reimbursed based on fees for diagnosis-related groups, reimbursements for residency training were calculated using cost reports in a base period, usually 1984. Line items related to salaries and benefits, and administrative expenses of residency programs were designated as direct costs of residency training. A per resident amount was calculated by dividing the total reported costs on these line items by the number of residents in the base period. Today, hospitals are reimbursed based on this per-resident amount, adjusted for inflation using CPI-U.

This reimbursement system therefore uses reported costs from two decades prior to the sample period of study. More importantly, the per resident amount may not reflect costs even in the base period because hospitals had little incentive to account for costs under the correct line item. Newhouse and Wilensky (2001) notes that the distinction between patient care costs from those incurred due to residency training is arbitrary and that variation in per-resident amounts may be driven by differences in hospital accounting practices or the use of volunteer faculty rather than real costs. In other words, whether a cost, say salaries paid to attending physicians, was accounted for in a line item later designated for direct costs can significantly influence reimbursement rates today.

These reimbursements are earmarked for costs of residency training and are positively associated with salaries paid by a program today (Table 2.8, Column 3). Reimbursement rates at competitor programs can therefore affect a program's salary offer because conversations with program directors suggest that salaries paid by competitors in a program's geographic area are used as benchmarks while setting their own salaries (Column 4).[22] I instrument using a weighted average of reimbursement rates of other teaching hospitals in the geographic area of a program. The instrument is defined as

$$r_j = \frac{\sum_{k \in G_j} fte_k \times rr_k}{\sum_{k \in G_j} fte_k}, \tag{2.7}$$

where rr_k and fte_k are the reimbursement rate and number of full-time equivalent residents at program k's primary hospital in the base period, and G_j are the hospitals in program j's geographic area other than j's primary hospital. I base the geographic definitions on Medicare's physician fee schedule, i.e. the MSA of the hospital or the rest of state if the hospital is not in an MSA. If less than three other competitors are in this area, define G_j to be the census division.[23]

Consistent with the theory for the instrument's effect on salaries, Column (5) shows that competitor reimbursements are positively related to salaries. Estimated in levels rather than logs, this specification is analogous to the first stage in a two-stage least-squares method.[24] In Column (6), I test the theory that competitor reimbursements affect salaries only through competitor salaries. Relative to column (5), controlling for the lagged average competitor

[22]Conversations with Dr. Weinstein, Vice President for GME at Partners Healthcare, suggest that salaries at residency programs sponsored by Partners Healthcare are aimed to be competitive with those at other programs in the Northeast and in Boston, by looking at market data from two publicly available sources (the COTH Survey and New England/Boston Teaching Hospital Survey).

[23]Additional details on Medicare's reimbursement scheme and the construction of the instrument are in Appendix B.7.

[24]Figure B.1 depicts this first stage visually. A strong increasing relationship between salary and competitor reimbursements is noticable. Clustered at the program level, the first stage F-statistic for the coefficient on the instrument is 37.6. Since the control function approach is based on assuming independence rather than mean independence, I test for heteroskedasticity in the residuals from the first stage. I could not reject the hypothesis that the residual is homoskedastic at the 90% confidence level for any individual year of data using either the tests proposed by Breusch and Pagan (1979) or by White (1980). Figure B.2 presents a scatter plot of the salary distribution against fitted values. The plot shows little evidence of heteroskedasticity.

salaries reduces the estimated effect of competitor reimbursements by an order of magnitude and results in a statistically insignificant effect.

The key assumption for validity of the instrument is that the program unobservable ξ_{jt} is conditionally independent of competitor reimbursement rates, given program characteristics and a program's own reimbursement rate, which is included in z_{jt} for specifications using the instrument. This assumption is satisfied if variation in reimbursement rates is driven by an arbitrary classification of costs by hospitals in 1984 or if past costs of competitors are not related to residents' preferences during the sample period. The primary threat is that reported per residents costs are correlated with persistent geographic factors. To some extent, this concern is mitigated by controlling for a program's own reimbursement rate. Reassuringly, Column (7) in Table 2.8 shows that the impact of competitor reimbursement rates on a program's salary changes by less than the standard error in the estimates upon including location characteristics such as median age, household income, crime rates, college population and total population.[25] Another concern is the possibility that programs respond to the reimbursement rates of competitors by engaging in endogenous investment. A comparison of estimates from Columns (2) and (5) shows little evidence of sensitivity of the coefficients on program characteristics (NIH, beds, Case Mix Index) to the inclusion of reimbursement rate variables.

2.6 Estimation

This section defines the estimator, the moments used in estimation, the simulation technique and a parametric bootstrap used for inference.

[25] Strictly speaking, the exclusion restriction requires that the instrument is not strongly correlated with factors that may determine choices of residents. Appendix B.7 shows that excluded location characteristics do not explain much variation in addition to controls included in the model although a formal test of exogeneity can be rejected.

2.6.1 Method of Simulated Moments

The estimation proceeds in two stages when the control function is employed. I first estimate the control variable ν_{jt} from equation (2.4) using OLS to construct the residual

$$\hat{\nu}_{jt} = w_{jt} - z_{jt}\hat{\gamma} - r_{jt}\hat{\tau}. \tag{2.8}$$

Replacing this estimate in equation (2.6), we get

$$u_{ijt} \approx z_{ijt}\beta_i^z + w_{jt}\beta_i^w + \kappa\hat{\nu}_{jt} + \sigma\zeta_{jt}, \tag{2.9}$$

where the approximation is up to estimation error in ν_{jt}. The estimation of parameters determining the human capital index of residents and their preferences over residents proceeds by treating $\hat{\nu}_{jt}$ like any other exogenous observable program characteristic. The error due to using $\hat{\nu}_{jt}$ instead of ν_{jt}, however, affects the calculation of standard errors. The first stage is not necessary in the model treating salaries as exogenous.

The distribution of preferences of residents and human capital can be determined as a function of observable characteristics of both sides and the parameter of the model, θ collected from equations (2.6), (2.2) and (3.1). The second stage of the estimation uses a simulated method of moments estimator (McFadden, 1989; Pakes and Pollard, 1989) to estimate the true parameter θ_0. The estimate $\hat{\theta}_{MSM}$ minimizes a simulated criterion function

$$\left\|\hat{m} - \hat{m}^S(\theta)\right\|_W^2 = \left(\hat{m} - \hat{m}^S(\theta)\right)' W \left(\hat{m} - \hat{m}^S(\theta)\right), \tag{2.10}$$

where $\hat{m}$ is a set of moments constructed using the matches observed in the sample, $\hat{m}^S(\theta)$ is the average of moments constructed from S simulations of matches in the economy, and W is a matrix of weights described in Section 2.6.4. Additional details on the estimator and the optimization algorithm are in Appendix B.1.[26]

2.6.2 Moments

The vector $\hat{m}$ consists of sample analogs of three sets of moments, stacked for each market and then averaged across markets. The simulated counterparts $\hat{m}^S(\theta)$ are computed

[26]The objective function in the specifications estimated have local minima, and is discontinuous due to the use of simulation. I use three starts of the genetic algorithm, which is a derivative-free global stochastic optimization procedure, followed by local searches using the subplex algorithm. Details are in Appendix B.1.

identically, but averaged across the simulations and markets. Mathematical expressions for the population versions and other details are in Appendix B.1.1.

For the match μ_t observed in market t, the set of moments are given by

1. Moments of the joint distribution of observable characteristics of residents and programs as given by the matches:
$$\hat{m}_{t,ov} = \frac{1}{N_t} \sum_{i \in \mathcal{N}_t} 1 \left\{ \mu_t \left(i \right) = j \right\} x_i z_{jt}. \tag{2.11}$$

2. The within-program variance of resident observables. For each scalar $x_{1,i}$:
$$\hat{m}_{t,w} = \frac{1}{N_t} \sum_{i \in \mathcal{N}_t} \left(x_{1,i} - \frac{1}{\left| \mu_t^{-1} \left(\mu_t \left(i \right) \right) \right|} \sum_{i' \in \mu_t^{-1} \left(\mu_t (i) \right)} x_{1,i'} \right)^2 . \tag{2.12}$$

3. The covariance between resident characteristics and the average characteristics of a resident's peers. For every pair of scalars $x_{1,i}$ and $x_{2,i}$:
$$\hat{m}_{t,p} = \frac{1}{N_t} \sum_{i \in \mathcal{N}_t} x_{1,i} \frac{1}{\left| \mu_t^{-1} \left(\mu_t \left(i \right) \right) \right| - 1} \sum_{i' \in \mu_t^{-1} \left(\mu_t (i) \right) \setminus \{i\}} x_{2,i'} . \tag{2.13}$$

The first set of moments include the covariances between program and resident characteristics. These moments are the basis of the regression coefficients presented in Tables 2.3 and 2.4. They quantify the degree of assortativity between resident and program characteristics observed in the data. I also include the probability that a resident is matched to a program located in the same state as her state of birth, or the same state as her medical school state.

The second and third set of moments take advantage of the many-to-one matching nature of the market.[27] Section 2.4.2 presents summaries of these moments from the data. The moments cannot be constructed in one-to-one matching markets, such as the marriage market, but are crucial to identify even the simpler double-vertical model. Since these moments extract information from within a peer group, they effectively control for both observable and unobservable program characteristics.[28]

---

[27] Alternatively, one could combine moments of type 1 and 2 to include all entries in the within program covariance of characteristics.

[28] Note that the number of moments suggested increases rapidly as more characteristics are included in

2.6.3 Simulating a Match

Under the parametric assumptions made on ζ_{jt}, ε_i, and η_i in Section 2.3, for a given parameter vector θ, a unique pairwise stable match exists and can be simulated. Because residents only participate in one market, matches of different markets can be simulated independently. For simplicity, I describe the procedure for only one market and omit the market subscript t. For a draw of the unobservables $\{\varepsilon_{is},\ \eta_{is}\}_{i=1}^{N}$ and $\{\zeta_{js}\}_{j=1}^{J}$ indexed by s, calculate

$$h_{is} = x_i\alpha + \varepsilon_{is}, \tag{2.14}$$

and the indirect utilities $\{u_{ijs}\}_{i,j}$. The indirect utilities determine the program resident i picks from any choice set.

Begin by sorting the residents in order of their simulated human capital, $\{h_{is}\}_{i=1}^{N}$, and let $i^{(k)}$ be the identity of the resident with the k-th highest human capital.

- *Step* 1 : Resident $i^{(1)}$ picks her favorite program. Set her simulated match, $\mu_s\left(i^{(1)}\right)$, to this program and compute $J^{(1)}$, the set of programs with unfilled positions after $i^{(1)}$ is assigned.

- *Step* $k > 1$: Let $J^{(k-1)}$ be the set of programs with unfilled positions after resident $i^{(k-1)}$ has been assigned. Set $\mu_s\left(i^{(k)}\right)$ to the program in $J^{(k-1)}$ most desired by $i^{(k)}$.

The simulated match μ_s can be used to calculate moments using equations (2.11) to (2.13). The optimization routine keeps a fixed set of simulation draws of unobservable characteristics for computing moments at different values of θ.

A model with preference heterogeneity on both sides requires a computationally more complex simulation method, such as the Gale and Shapley (1962) deferred acceptance algorithm (DAA), to compute a particular pairwise stable match. In the DAA, each applicant

the preference models. If the covariance between each observed characteristic of the resident and of the program are included in the first set of moments, the number of moments is at least the product of the number of characteristics of each side. On the other hand, the number of parameters is the sum of the number of characteristics. This relative growth can create difficulties when estimating models with a very rich set of characteristics.

simultaneously applies to her most favored program that has not yet rejected her. A set of applications are held at each stage while others are rejected and assignments are made final only when no further applications are rejected. This temporary nature of held applications and the need to compute a preferred program for all applications at each stage significantly increases the computational burden for a market with many participants such as the one studied in this paper.[29]

2.6.4 Econometric Issues

In a data environment with many independent and identically distributed matching markets, the sample moments and their simulated counterparts *across* markets can be seen as iid random variables. Well known limit theorems could be used to understand the asymptotic properties of a simulation based estimator (McFadden, 1989; Pakes and Pollard, 1989). The data for this study are taken from eight academic years, making asymptotic approximations based on data from many markets undesirable. Within each market, the equilibrium match of agents are interdependent through both observed and unobserved characteristics of other agents in the market. For this reason, modelling the data generating process as independently sampled matches is unappealing as well.

Instead, I consider a data generating process in which the size of the market grows rather than the number of markets. The family medicine residency market has about 430 programs and 3,000 residents participating each year. Similar facts motivated theoretical work on the structure of the set of stable matches and incentives of agents as the market grows in size (Kojima and Pathak, 2009)

Agarwal and Diamond (2013) studies the properties of the estimator for the double-vertical model in a single market for a data generating process in which the number of programs and residents increases. For each program, j, the capacity is drawn from the distribution F_c, with support on the natural numbers less than $\bar{c}$. They study the case where the total number of positions $C_{tot} = \sum_j c_j$ is equal to the number of residents N. Under

[29]Even with an insertion sort, a relatively inefficient sorting algorithm, the computational complexity of the algorithm used here is $O\left(n^2\right)$ whereas if preferences were heterogenous on both sides, a simulation to calculate the resident optimal match using deferred acceptance algorithm would have a computational complexity of $O\left(n^3\right)$.

these asymptotics, the number of market participants on each side grows at a stochastically proportional rate. The observed data is a pairwise stable match for N residents and J programs with characteristics (x_i, ε_i) and (z_{jt}, ξ_{jt}) drawn from their respective population distributions. Such data can be viewed as a joint distribution of observable characteristics of programs and residents, with information also on each resident's peer group in the program. The challenge in obtaining asymptotic theory arises precisely from the dependence of matches on the entire sample of observed characteristics. Similar challenges arise in the literature on network formation models (see Christakis et al., 2010; Kolaczyk, 2009). Monte Carlo evidence suggests that in a more general model like the one estimated in this paper, the root mean square error in parameter estimates decreases with the sample size.

Calculating Standard Errors

An additional challenge arises for constructing confidence sets for the estimated parameter because of interdependence of matches, and because bootstrapping the estimator directly is computationally prohibitive. The covariance of the moments is estimated using a parametric bootstrap to account for the dependence of matches across residents. With this estimate, I approximate the error in the estimated parameter using a delta method that is commonly used in simulated estimators (Gourieroux and Monfort, 1997):

$$\hat{\Sigma} = \left(\hat{\Gamma}' W \hat{\Gamma}\right)^{-1} \hat{\Gamma}' W \left(\hat{V} + \frac{1}{S}\hat{V}^S\right) W' \hat{\Gamma} \left(\hat{\Gamma}' W \hat{\Gamma}\right)^{-1}, \tag{2.15}$$

where $\hat{\Gamma}$ is the gradient of the moments with respect to θ evaluated at $\hat{\theta}_{MSM}$ using two-sided finite-difference derivatives; W is the weight matrix used in estimation; $\hat{V}$ is an estimate of the covariance of the moments at $\hat{\theta}_{MSM}$; S is the number of simulations and $\hat{V}^S$ is an estimate of the simulation error in the moments at $\hat{\theta}_{MSM}$.

In this section, I describe the choice of W and outline the parametric bootstrap used to estimate $\hat{V}$ for the simpler case with $N = C_{tot}$ and exogenous salaries. Appendix B.1 provides additional details on estimating $\hat{\Sigma}$. The bootstrap mimics the data generating process described earlier. Three basic steps are used for each bootstrap iteration $b \in \{1, \ldots, B\}$:

1. Generate a bootstrap sample of programs $\{z_{j,b},\ c_{j,b}\}_{j=1}^{J}$ by drawing from the empirical distribution $\hat{F}_{Z,C}$ with replacement. Calculate $C_{tot,b} = \sum_j c_{j,b}$.

2. Generate a bootstrap sample of residents $\{x_{i,b}\}_{i=1}^{C_{tot,b}}$ from $\hat{F}_X$, with replacement.

3. Simulate the unobservables $(\varepsilon_{i,b},\ \nu_{i,b},\ \xi_{jt,b})$ to compute $\{h_{i,b}\}_{b=1}^{C_{tot,b}}$ and $\{u_{i,j,b}\}_{i,j}$ at $\hat{\theta}_{MSM}$. Calculate the stable match μ_b for bootstrap b and corresponding moments $\hat{m}^b$.

The variance of $\hat{m}^b$ is the estimate for $\hat{V}$ used to compute $\hat{\Sigma}$. Monte Carlo evidence suggests that the procedure yields confidence sets with close to the correct size. The model using the control function correction has an additional step in this bootstrap to account for uncertainty in estimating $\hat{\nu}_{jt}$, also described in Appendix B.1.

Finally, the weight matrix in estimation is obtained from bootstrapping directly from the joint distribution of matches observed in the data. A bootstrap sample of matches $\{\mu_b\}_{b=1}^{B}$ is generated by sampling, with replacement, J programs and along with their matched residents. The moments from these matches are computed and the inverse of the covariance is used as the positive definite weight matrix, W. The procedure does not require a first step optimization and does not need to converge to $\hat{V}^{-1}$.

2.7 Empirical Specifications and Results

I present estimates from three models. The first model has the richest form of preferences as it allows for unobserved heterogeneity in preferences via normally distributed random coefficients on Case Mix Index, NIH Funds of major medical school affiliates and the number of beds. It also allows for heterogeneity in taste for program location based on a resident's birth location and medical school location. I use a second model that does not include random coefficients on Case Mix, NIH Funds or beds to assess the importance of unobserved preference heterogeneity. These two models treat salaries as exogenous. The final model modifies the second model to addresses the potential endogeneity in salaries using the instrument described in Section 2.5.2. This specification includes a program's own reimbursement rate in addition to characteristics included in the other models.

Estimates of residents' preferences for programs presented in the next section are translated into dollar equivalents for a select set of program characteristics. I also present the willingness to pay by categories of programs. These are the most economically relevant statistics obtained from preference estimates. Appendix B.2 briefly discusses the underlying parameters, which are not economically intuitive, and robustness using estimates from additional models.

2.7.1 Preference Estimates

Panel A.1 of Table 2.7 presents the estimated preferences for programs in salary equivalent terms. Comparing specifications (1) and (2), the estimated value of a one standard deviation higher Case Mix Index at an otherwise identical program is about \$2,500 to \$5,000 in annual salary for a typical resident. Likewise, residents are willing to pay for programs at larger hospitals as measured by beds, and for programs with better NIH funded affiliates. The estimates from specification (1) suggest a substantial degree of preference heterogeneity for these characteristics as well. The additional heterogeneity in preferences relative to specification (2) results in a shift in the mean willingness to pay for NIH funding of major affiliates, the Case Mix Index, and beds, but not whether they are desirable or not.

Panel A.2 presents estimates of preferences for program types and heterogeneity in preferences for program location. Both specifications (1) and (2) estimate that, *ceteris paribus*, rural programs are preferable to urban programs. This result is consistent with the reduced form evidence presented in Section 2.2, which shows a positive though statistically insignificant association between resident quality and rural programs, and that rural programs do not have a significantly larger fraction of unfilled positions than urban programs. Because rural programs tend to be associated with smaller hospitals and medical school affiliates with lower NIH funding, these estimates do not necessarily imply that rural programs are preferred to urban programs. The next section presents the willingness to pay by program categories and shows that overall, rural programs are less preferred to urban programs.

Estimates from both specifications also suggest that residents prefer programs in their state of birth or in the same state as their medical school. For instance, estimates from specification (1) imply that a typical resident is willing to forgo about \$10,000 in salary to match at a program in the same state as their medical school. Although rural born residents prefer rural programs more than other residents, they prefer rural programs at a monetary

Table 2.7: Preference Estimates

	Full Heterogeneity (1)	Geographic Heterogeneity (2)	Geo. Het. w/ Wage Instrument (3)
Panel A.1: Preference for Programs (units of std. dev)			
Case Mix Index			
Coeff	4,792	2,320	6,088
	(1,624)	(1,265)	(1,542)
Sigma RC	4,503		
	(1,037)		
Log NIH Fund (Major)			
Coeff	491	6,499	4,402
	(1,651)	(2,041)	(1,333)
Sigma RC	5,498		
	(1,234)		
Log Beds			
Coeff	6,900	3,528	8,837
	(2,207)	(1,259)	(1,936)
Sigma RC	11,107		
	(2,073)		
Log NIH Fund (Minor)	4,993	5,560	7,620
	(1,558)	(1,511)	(1,821)
Panel A.2: Preference for Programs			
Rural Program	7,327	5,611	17,314
	(3,492)	(3,555)	(4,938)
University Based Program	15,786	11,080	25,130
	(3,982)	(5,393)	(7,088)
Community/University Program	-5,001	-2,217	-7,507
	(2,016)	(1,589)	(2,233)
Medical School State	9,820	2,302	4,529
	(1,998)	(687)	(910)
Birth State	6,342	1,320	2,451
	(1,308)	(411)	(497)
Rural Birth x Rural Program	1,189	109	233
	(466)	(113)	(102)

(cont'd...)

Table 2.7: Preference Estimates (cont'd)

	Full Heterogeneity (1)	Geographic Heterogeneity (2)	Geo. Het. w/ Wage Instrument (3)
Panel B: Human Capital			
Log NIH Fund (MD)	0.1153	0.1269	0.0941
	(0.0164)	(0.0139)	(0.0131)
Median MCAT (MD)	0.0814	0.0666	0.0413
	(0.0070)	(0.0038)	(0.0030)
US Born (Foreign Grad)	0.1503	-0.2470	0.2927
	(0.1021)	(0.0801)	(0.0705)
Sigma (DO)	0.8845	0.7944	0.7275
	(0.0359)	(0.0285)	(0.0292)
Sigma (Foreign)	3.6190	3.0709	2.8215
	(0.1469)	(0.1102)	(0.1131)

Notes: Detailed estimates and other models using instruments in Table B.1. Results from Panel A estimates monetized in dollars (normalize wage coefficient to 1). Panel A.1 presents the dollar equivalent for a 1 standard deviation change in a program characteristic. All columns include median rent in county, Medicare wage index, indicator for zero NIH funding of major associates and for minor associates. Column (4) includes own reimbursement rates and the control variable. All specifications normalize the mean utility from a program with zeros on all characteristics to 0. In all specifications, the variance of unobservable determinants of the human capital index of MD graduates is normalized to 1. All specifications normalize the mean human capital index of residents with zeros for all characteristics to 0 and include medical school type dummies. Point estimates using 1000 simulation draws. Standard errors in parenthesis. Optimization and estimation details described in an appendix.

equivalent of under $1,200. The estimated willingness to pay for these factors is smaller in specification (2) although the relative importance for the different dimensions is similar.

Panel B presents parameter estimates for the distribution of human capital, which determines ordinal rankings between residents. All specifications yield similar coefficients on the various resident characteristics and estimate that the unobservable determinants of human capital have larger variances for residents with foreign degrees. The estimated difference between a US born foreign medical graduate and foreign graduates from other countries is an order of magnitude smaller than the standard deviation of unobservable determinants of human capital.

Estimates with Instruments

As compared to estimates from specification (2), which treats salaries as exogenous, the estimated willingness to pay for program characteristics is generally larger in specification (3). The estimates for NIH funding of Major Medical school affiliates is the only exception. The increase in the estimated willingness to pay in specification (3) is driven by a fall in the coefficient on salaries but similar coefficient estimates for the other program characteristics. Appendix B.2 discusses results from the instrumented version of specification (1), which also leads to a decrease in the coefficient on salaries and little change in estimates for other coefficients. This specification results in a small, positive coefficient on salaries that is not statistically significant and implies an implausibly large willingness to pay for better programs. The qualitative effect of including the wage instrument on parameter estimates indicates that, if anything, treating salaries as exogenous may lead to an understated willingness to pay for more desirable programs. I interpret the magnitudes with caution given the lack of robustness, which is likely a consequence of the limited salary variation in the data.[30] Aside for controlled geographic covariates such as rent and wage index, estimates in Column (2) of Table 2.8 do not show strong evidence of substantial correlation of salaries with program characteristics. My preferred approach is to focus on results from specification (1) for most counterfactual results and discuss the effect of possible positive

[30]The objective function for specifications using salary instruments is fairly flat along different combinations of coefficients on the wage and control variables.

Table 2.8: Wage Regressions

	Dependent Variable: Log First Year Salary						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log Rent	0.0266*	-0.0373**	-0.0379**	0.0179	-0.0378**	0.0172	-0.0306
	(0.0151)	(0.0177)	(0.0175)	(0.0140)	(0.0160)	(0.0143)	(0.0230)
Rural Program	0.0032	0.0065	0.0110	0.0103	0.0104	0.0103	0.0055
	(0.0079)	(0.0081)	(0.0080)	(0.0071)	(0.0076)	(0.0069)	(0.0079)
Log Wage Index		0.1366***	0.1182***	-0.0152	0.0806***	-0.0167	0.0809***
		(0.0307)	(0.0302)	(0.0262)	(0.0287)	(0.0263)	(0.0290)
Log NIH Fund (Major)		0.0024	0.0023	0.0062***	0.0034	0.0062***	0.0024
		(0.0027)	(0.0026)	(0.0021)	(0.0025)	(0.0021)	(0.0024)
Log NIH Fund (Minor)		-0.0060*	-0.0047	-0.0005	-0.0040	-0.0005	-0.0041
		(0.0032)	(0.0032)	(0.0029)	(0.0031)	(0.0029)	(0.0031)
Log # Beds		0.0087*	0.0086*	0.0012	0.0064	0.0010	0.0108**
		(0.0046)	(0.0045)	(0.0036)	(0.0041)	(0.0036)	(0.0043)
Log Case-Mix Index		-0.0108	-0.0046	0.0051	-0.0038	0.0056	-0.0065
		(0.0195)	(0.0195)	(0.0151)	(0.0190)	(0.0152)	(0.0191)
Log Reimbursement			0.0227***		0.0064	-0.0002	0.0050
			(0.0077)		(0.0076)	(0.0063)	(0.0070)
Log Competitor Salary (Lagged)				0.8779***		0.8651***	
				(0.0542)		(0.0683)	
Log Competitor Reimbursement					0.0968***	0.0090	0.0847***
					(0.0170)	(0.0170)	(0.0178)
Location characteristics							Y
Observations	3,418	3,418	3,418	2,997	3,418	2,997	3,418
R-squared	0.0062	0.0452	0.0640	0.3284	0.1226	0.3294	0.1811

Notes: Location characteristics include median age (county), log median household income (county), log total population (MSA/county), violent crime and property crime rates, dummies for no data in that radius and log college share (MSA/rest of state). Columns (2-7) include dummies for programs with no NIH funding at major affiliates, minor affiliates, and for missing Medicare ID for the primary institution. See data appendix for additional details on variable construction. Standard errors clustered at the program level.

bias in the salary coefficient using specification (3).

Distribution of Willingness to Pay

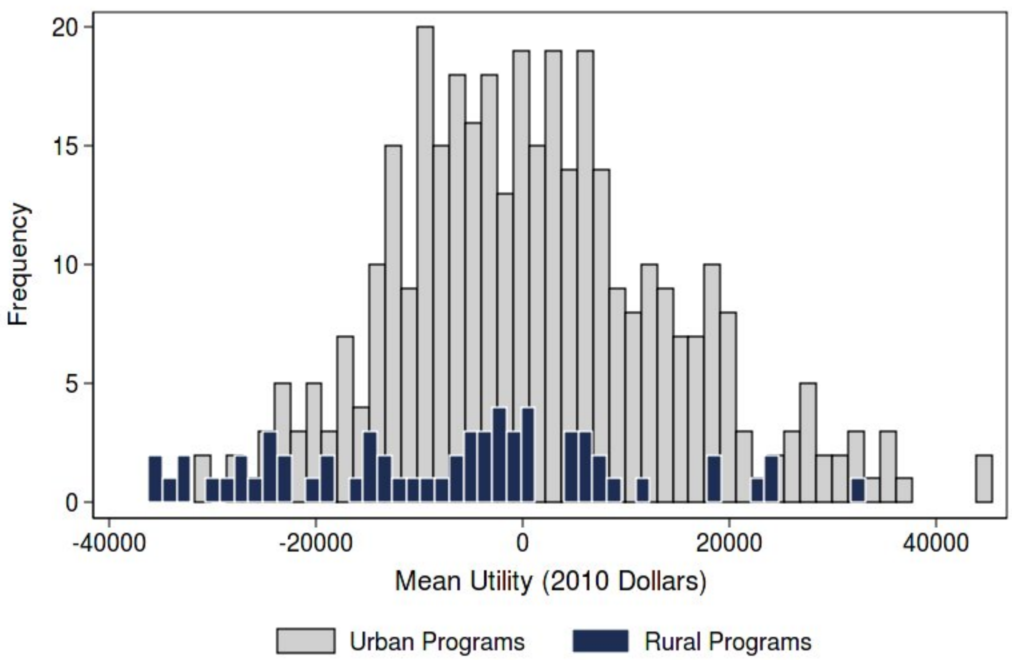
The distribution of willingness to pay for different programs is an important economic input for analyzing salaries under competitive wage bargaining and for evaluating the effect of financial incentives for rural training. Figure 2.2 plots the estimated distribution of utility (in dollars) across programs averaged over residents, net of salaries, for the 2010-2011 sample year as implied by specification (1). This sample will be used for all counterfactual exercises. Table 2.9 presents summary statistics of this distribution by categorizing programs into quartiles based on observed characteristics, and normalizing the mean across all programs to zero. I estimate a large willingness to pay for programs with a high Case Mix Index, at larger hospitals and in counties with larger programs. A typical resident is willing to accept a \$5,000 to \$9,500 lower salary at the average urban program instead of a training in a rural location. At under \$1,200, the estimated additional preference of rural born residents for a rural program is not sufficient to overturn the mean distaste for training in rural programs. The finding that the typical rural hospital is not substantially less attractive than their urban counterparts is consistent with conclusions of Rosenblatt et al. (2006). Using surveys of program directors, they find that residents matched at rural programs and the number of applications per position are similar to those in urban programs.

Specifications (1) and (2) estimate the standard deviation in utility across residents and programs of varying characteristics to be between \$14,000 and \$22,000. This measure doubles from \$14,000, but is imprecisely estimated, when Specification (2) is modified to account for endogeneity in salaries. While differences in the quality of training provided by a program is likely the primary driver of willingness to pay for different programs, as evidenced by tastes for geographically nearby programs, there may be some contemporaneous value for desirable amenities. At first glance, the estimated standard deviation in willingness to pay for programs may seem large with respect to the observed variation in salaries (about \$3,200). However, the ideal comparison is with the distribution of training value added in terms of future income across residency programs, which is likely much larger. Such a comparison is not possible given the available data.

Table 2.9: Estimated Utility Distribution in First-Year Salary Equivalent

	N	Full Heterogeneity (1) Stat	(s.e.)	Geographic Heterogeneity (2) Stat	(s.e.)	Geo. Het. w/ Wage Instrument (3) Stat	(s.e.)
Panel A: Means in Category							
Log Beds (Primary Inst)							
Lowest Quartile	107	-\$12,509	(3,290)	-\$5,691	(777)	-\$15,238	(4,647)
Second Quartile	107	-\$2,801	(758)	-\$3,693	(553)	-\$3,606	(1,212)
Third Quartile	107	\$3,823	(1,138)	-\$1,041	(320)	\$1,934	(1,108)
Highest Quartile	107	\$11,487	(2,877)	\$10,425	(1,327)	\$16,910	(4,831)
Case Mix Index							
Lowest Quartile	107	-\$10,397	(2,880)	-\$4,045	(674)	-\$10,556	(3,450)
Second Quartile	107	-\$3,764	(1,100)	-\$1,965	(436)	-\$5,162	(1,643)
Third Quartile	107	\$3,346	(1,179)	-\$1,518	(403)	\$669	(720)
Highest Quartile	107	\$10,815	(2,849)	\$7,528	(1,196)	\$15,050	(4,663)
Log NIH Fund (Major)							
Lowest Quartile	71	-\$5,190	(1,716)	-\$7,903	(1,064)	-\$15,032	(4,267)
Second Quartile	71	-\$3,712	(1,080)	-\$285	(390)	-\$8,095	(2,685)
Third Quartile	71	\$1,796	(963)	\$8,460	(1,274)	\$6,646	(2,021)
Highest Quartile	72	\$904	(1,535)	\$11,733	(1,736)	\$7,194	(2,368)
County Rent							
Lowest Quartile	106	-\$5,681	(1,580)	-\$6,745	(984)	-\$11,796	(3,549)
Second Quartile	107	-\$1,012	(541)	-\$964	(244)	-\$3,310	(1,077)
Third Quartile	99	\$1,984	(688)	\$1,715	(333)	\$2,942	(1,204)
Highest Quartile	116	\$4,431	(1,321)	\$5,589	(827)	\$11,321	(3,148)
Rural Program	63	-\$7,292	(3,101)	-\$4,692	(967)	-\$8,066	(4,044)
Urban Program	365	\$1,259	(535)	\$810	(167)	\$1,392	(698)
Overall Std. Dev.	428	\$21,937	(5,215)	\$14,088	(1,880)	\$28,578	(8,166)

Notes: Utilities net of salaries are monetized in dollars and normalized to an overall mean of zero. Statistics averages across residents from 100 simulation draws. Each simulation draws a parameter from a normal with mean $\hat{\theta}_{MSM}$ and variance $\hat{\Sigma}$, where $\hat{\Sigma}$ is estimated as described in Section 2.6.4. Statistics use the 2010-2011 sample.



Notes: Estimated distribution of mean utility (from observable components, net of salary) across programs monetized in terms of first year salary. Mean utility normalized to zero. Sample of programs from 2010-2011.

Figure 2.2: Estimated Distribution of Program Utility

2.7.2 Model Fit

In this section, I describe the in-sample and out-of-sample fit of estimates from specification (1). The fit of specifications (2) and (3) are qualitatively similar. The out-of-sample fit uses data from the 2011-2012 wave of the GME Census, which was only accessed after parameter estimates were computed.

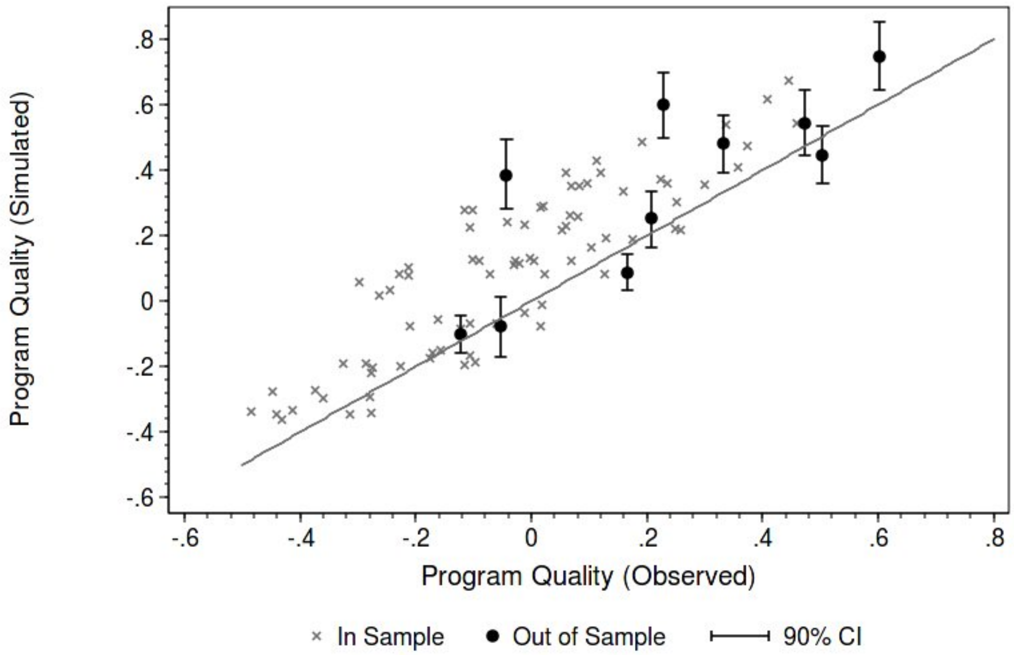
Estimates of the model only determine the probability that a resident with a given observable characteristic matches with a program with certain observables. The uncertainty in matches arises from unobservables of both the residents and the programs. Therefore, an assessment of fit must use statistics that average matches across groups of residents or programs.

For simplicity of exposition, I assess model fit using a single dimensional average quality of matched program for a group of residents with similar observable determinants of human capital. For each year t, I use the parameter estimates from the model to construct a quality index for each resident i and program j by computing $x_i\hat{\alpha}$ and $z_{jt}\hat{\beta}$ respectively. Then, I divide the residents into ten bins based on $x_i\hat{\alpha}$ and compute the mean quality of program with which residents from each bin are matched. Figure 2.3 presents a binned scatter plot of this mean quality of program as observed in the data and predicated by model simulations. Both the in-sample points and the out-of-sample points are close to the 45-degree line. The 90% confidence sets of the simulated means for several resident bins include the theoretical prediction. [31]

This fit of the model provides confidence that parametric restrictions on the model are not leading to poor predictions of the sorting patterns in the market. Therefore, I am comfortable using estimates as basis of counterfactual analysis.

---

[31] A more model-free assessment of fit using sorting regressions only on observed covariates is presented in Table B.2. One may also worry predicting sorting patterns is is mechanical because there is little change in the market composition across years. For counterfactuals directly impacting the composition of market participants, it can be important for the model to capture changes in sorting as a function of changes in the composition of the market. However, changes in the composition of the resident and program distribution are negligible, resulting in little available variation to test the model with such a fit.



Notes: To construct this scatterplot, I used model estimates from specification (1) to first obtain the predicted quality on observable dimensions of the residents and of the programs. Quality for the program is the "vertical component" $z_j\beta$ for the programs. The residents were binned into 10 categories, starting with Foreign graduates, US born foreign graduates and Osteopathic graduates and seven quantile bins for MD graduates. Resident bins are constructed from pooling the sample across all years. The seven MD bins are approximately equally sized, except for point masses at the cutoffs. The horizontal axis plots observed mean standardized quality of program that residents from each bin matched with. The vertical axis plots the model's predicted mean standardized quality of the program that a resident in each bin is matched with. An observation is defined at the bin-year level. Simulated means using the observed distribution of agent characteristics and 100 simulations of the unobserved characteristics. The 90% confidence set for the out-of-sample data is constructed from these 100 simulations.

Figure 2.3: Model Fit: Simulated vs. Observed Match Quality by Resident Bins

2.8 Application 1: Salary Competition

In 2002, a group of former residents brought on a class-action lawsuit under the Sherman Act against major medical associations in the United States and the NRMP. The plaintiffs alleged the medical match is an instrumental competitive restraint used by the residency programs to depress salaries.[32] By replacing a traditional market in which residents could use multiple offers to negotiate with programs, they argued that the NRMP "enabled employers to obtain resident physicians without such a bidding war, thereby artificially fixing, depressing, standardizing and stabilizing compensation and other terms of employment below competitive levels" (Jung et.al. v AAMC et.al., 2002). A brief prepared by Orley Ashenfelter on behalf of the plaintiffs argued that competitive outcomes in this market would yield wages close to the marginal product of labor, which was approximated using salaries of starting physicians, nurse practitioners, and physician assistants.[33] Physician assistants earned a median salary of $86,000 in 2010[34] as compared to about $47,000 for medical residents despite longer work hours.[35]

Recent papers have debated whether low salaries observed in this market are a results of the match. Using a stylized model, Bulow and Levin (2006) argue that salaries may be depressed in the match because residency programs face the risk that a higher salary may not necessarily result in a better resident. Kojima (2007) uses an example to show that this result is not robust in a many-to-one matching setting because of cross-subsidization across residents in a program. Empirical evidence in Niederle and Roth (2003a, 2009) suggests that medical fellowship salaries are not affected by the presence of a match, however, the study does not explain why fellowship salaries remain lower than salaries paid to other

[32] Jung et.al. v AAMC et.al. (2002) states that "The NRMP matching program has the purpose and effect of depressing, standardizing and stabilizing compensation and other terms of employment." After the lawsuit was filed, the Pension Funding and Equity Act of 2004 amended antitrust law to disallow evidence of participation in the medical match in antitrust cases. The lawsuit was dismissed following this amendment, overturning a previous opinion of the court upholding the price-fixing allegation.

[33] A redacted copy of the expert report submitted on behalf of the plaintiffs is available on request.

[34] Source: Bureau of Labor Studies.

[35] At 50 work-weeks a year and 80 hour a week, the cap imposed by the ACGME in 2003, a salary of $50,000 yields a wage rate for a medical resident of $12.50. A more generous estimate with 65 hours a week, 45 work-weeks a year and a salary of $60,000 yields a wage rate of $20.50.

health professionals.

The plaintiffs argued their case based on a classical economic model of homogeneous firms competing for the services of labor and free entry. However, such a perfect competition benchmark may not be a good approximation for an entry-level professional labor market. The data provide strong evidence that residents have preferences for characteristics of the program other than the wages and may, thus, reject a higher salary offer from a less desirable program. Further, barriers to entry by residency programs are high and capacity constraints are imposed by accreditation requirements. A program must therefore consider the option value of hiring a substitute resident when confronted with a competing salary offer. High quality programs may be particularly able to find other residents willing to work for low salaries. Conversely, highly skilled residents are scarce and they may be able to bargain for higher salaries. It is essential to consider these incentives in order to predict outcomes under competitive salary bargaining.

I model a "traditional" market using a competitive equilibrium, which is described by a vector of worker-firm specific salaries and an assignment such that each worker and firm demands precisely the prescribed assignment. Shapley and Shubik (1971) show that competitive equilibria correspond to core allocations and satisfy two conditions. First, allocations must be individually rational for both workers and firms. Second, it must be that at the going salaries no worker-firm pair would prefer to break the allocation to form a (different) match at renegotiated salaries. This latter requirement ensures that further negotiations cannot be mutually beneficial. Kelso and Crawford (1982) show that competitive equilibria can result from a salary adjustment process in which the salaries of residents with multiple offers are sequentially increased until the market clears. The process embodies the "bidding war" plaintiffs suggest would arise in a "traditional" market. Crawford (2008) proposed a redesign of the residency match based on the salary adjustment process with the aim of increasing the flexibility of salaries in the residency market and implementing a competitive equilibrium outcome.

I first develop a stylized model to derive the dependence of competitive equilibrium salaries on both the willingness to pay for programs and the production technology of residency programs. For counterfactual simulations, I adopt an approach that does not rely on knowing the production technology of resident-program pairs because data on residency program output is not available. Instead of calculating equilibrium salaries, I use the estimates of only the residents' preferences to calculate an equilibrium markdown from output net of training costs, called the implicit tuition. Loosely speaking, my calculation acts as if the output produced by a program-resident pair accrues entirely to residents. The illustrative model shows that the approach is likely to understate the equilibrium markdown in salaries since programs do not earn any infra-marginal productive rents due to their own productivity. The theoretical model is also used to describe differences with related models of on-the-job training or salary setting with non-pecuniary amenities.

2.8.1 An Illustrative Assignment Model

I generalize the model of the residency market in Bulow and Levin (2006) which assumes that residents take the highest salary offer. I allow resident preferences to depend on program quality in addition to salaries, and use a more flexible production function than Bulow and Levin (2006).

Consider an economy with N residents and programs in which each program may hire only one resident. Resident i has a human capital index, $h_i \in [0, \infty)$, and program j has a quality of training index, $q_j \in [0, \infty)$. To focus on salary bargaining, the training quality of programs are held exogenous. Without loss of generality, index the residents and programs so that $h_i \geq h_{i-1}$, $q_j \geq q_{j-1}$, and q_1 and h_1 are normalized to zero.

Residents have homogenous, quasi-linear preferences for the quality of program, $u(q, w) = aq + w$ with $a \geq 0$. The value, net of variable training costs, to a program of quality q of employing a resident with human capital index h is $f(h, q)$ where f_h, f_q, $f_{hq} > 0$ and $f(0, 0)$ is normalized to 0.[36] A program's profit from hiring resident h at salary level w is $f(h, q) - w$. I assume that an allocation is individually rational for a resident if $u(q, w) \geq 0$, and for a program if $f(h, q) - w \geq 0$.

A competitive equilibrium assignment maximizes total surplus. In this model, the unique equilibrium is characterized by positive assortative matching and full employment. Hence, in equilibrium, resident k is matched with program k and is paid a possibly nega-

[36] A complementary production technology is commonly assumed for studying on-the-job training (Becker, 1975, pp 34) or sorting in matching markets (Becker, 1973; Teulings, 1995).

tive wage w_k. The vector of equilibrium wages is determined by the individual rationality constraints and the constraint

$$f\left(h_k, q_k\right) - w_k \geq f\left(h_i, q_k\right) - w_i + a\left(q_k - q_i\right). \tag{2.16}$$

This constraint on w_k requires that the profit of program k by hiring resident k must be weakly greater than the profit from hiring resident i. At the going salaries, it is incentive compatible for resident i to accept an offer from program k only if the wage is at least $w_i - a\left(q_k - q_i\right)$.

There is a range of wages that are a part of a competitive equilibrium. Shapley and Shubik (1971) shows that there exists an equilibrium that is weakly preferred by all residents to all other equilibria, and another that is preferred by all programs. Appendix B.3.1 characterizes the entire set of equilibria, and derives the expression for wages at these two extremal outcomes. Since the plaintiffs alleged that salaries are currently much lower than in a bargaining process, I focus on the worker-optimal equilibrium which has higher salaries for every worker than any other equilibrium. This outcome is unanimously preferred by all residents to other competitive equilibria. The wage of resident k in the worker optimal equilibrium is given by

$$w_k = -a q_k + \sum_{i=2}^{k} \left[f\left(h_i, q_i\right) - f\left(h_{i-1}, q_i\right) \right]. \tag{2.17}$$

Resident 1 receives her product of labor $f\left(h_1, q_1\right)$ (normalized to 0), the maximum her employer is willing to pay. For resident 2, the first term aq_2 represents an implicit price for the difference in the value of training received by her compared to that of program 1 (with $q_1 = 0$). If a resident were to use a wage offer of w by program 1 in a negotiation with program 2, the resident would accept a counter offer of $w - aq_2$. The second term in this resident's wage, $f\left(h_2, q_2\right) - f\left(h_1, q_2\right)$, is program 2's maximum willingness to pay for the difference in productivity of residents 1 and 2, which accrues entirely to the resident in the worker-optimal equilibrium. The sum of these two terms measures the impact of the outside option of each party on the wage negotiation determining w_2. For $k > 2$, these (local) differences in the productivity of residents add up across lower matches to form the equilibrium wage.

Implicit Tuition

The implicit price for training at firm k, given by aq_k, is based on the preferences for training at a program rather than the cost of training. In models of general training that use a perfect competition framework, such as Rosen (1972) and Becker (1975), the implicit price is the marginal cost of training alone because free entry prevents firms from earning rents due to their quality.[37] When entry barriers are large due to fixed costs or restrictions from accreditation requirements, firms can earn additional profits due to their quality. I argue that ruling out entry is appropriate because of accreditation requirements and to focus on wage bargaining. Equation (2.17) shows that under these assumptions, program k can levy the implicit tuition aq_k on residents. This implicit tuition results from a force similar to compensating differentials (Rosen, 1987), but allows for heterogeneity in resident skill. Equilibrium salaries are the sum of the implicit tuition and a split of the value f produced by a resident program pair.

As mentioned earlier, the data does not allow us to determine f. I calculate the implicit tuition using residents' preferences alone in order to evaluate whether a gap between f and equilibrium salaries exists as a result of market fundamentals. The next result shows that the implicit tuition bounds the markdown in salaries from below. Under free entry by firms, salaries would be equal to f because any profits earned by firms would be competed away.

Proposition 2.1. *For all production functions f with $f_h,\ f_q,\ f_{hq} \geq 0$, the profits of the firm k is bounded below by aq_k in any competitive equilibrium.*

Proof. Corollary to Proposition B.3 stated and proved in Appendix B.3.2. □

Hence, the implicit tuition aq_k is a markdown in salaries that is independent of the output. If residents have a strong preference for program quality, this implicit tuition will be large and salaries in any competitive equilibrium are well below the product $f(h_k, q_k)$.

[37] Viewing $f(h, q)$ as output net of costs of training, a constant training cost across residents and programs would shift the wage schedule down by that constant. As can be seen from equation (2.17), training costs that depend on program quality, but not the quality of the resident do not affect equilibrium salaries as long as f_q remains positive. Also note that the implicit price aq_k does not depend on the number of residents and programs N, which could be very large, or the distribution of program quality. Intuitively, the important difference overturning results from perfect competition is that the number of firms competing for a fixed set of workers is not disproportionately large.

To interpret the implicit tuition as a lower bound for salary markdowns, consider two particular limiting cases for the production function. If $f(h, q)$ depends only on h so that the value of a resident, denoted $\bar{f}(h)$, does not vary across programs, the worker-optimal salaries are given by

$$w_k = \bar{f}(h_k) - aq_k. \tag{2.18}$$

Under this production function, the resident is the full claimant of the value of her labor and salaries equal her product net of the implicit tuition. Residents are able to engage programs in a bidding war until their salary equals the output less the implicit tuition because all programs value resident k at $\bar{f}(h_k)$.

On the other hand, if $f(h, q)$ depends only on q so that all residents produce $\underline{f}(q)$, irrespective of their human capital, the worker-optimal salaries are

$$w_k = -aq_k. \tag{2.19}$$

In this case, the program does not share the product $\underline{f}(q_k)$ with the resident since any two residents are equally productive at the program. The resident still pays an implicit tuition for training.[38]

The production function directly influences competitive salaries but Proposition 2.1 shows that in all cases resident k pays the implicit tuition aq_k. Equilibrium wages given in equations (2.18) and (2.19) highlight that the side of the market that owns the factor determining differences in f is compensated for their productivity in a competitive equilibrium. Residents are compensated for their skill only if human capital is an important determinant of f. For this reason, using a production function of the form $\bar{f}(h)$ results in a markdown in salaries from f that is only due to the implicit tuition.

This interpretation highlights a key difference from results derived using models with many firms competing for labor with free entry. In those models, one expects all the product to accrue to the workers because firms enter the market to bid for labor services until a zero profit condition is met. High compensation for residents is a result of free entry rather than negotiations between a fixed set of agents.

[38] In order to ensure that the match is assortative in these limiting cases, I assume that if a program (resident) has two equally attractive offers, the tie in favor of the resident (program) with the higher human capital (quality).

2.8.2 Generalizing the Implicit Tuition

The expression for the implicit tuition derived above relied on the assumption that residents have homogeneous preferences for program quality. For this reason, the results from the illustrative model do not speak to competitive outcomes in a model with heterogenous preferences. This section generalizes the definition of implicit tuition to make it applicable to the model defined in Section 2.3.

Notice that the profit earned by program k in a worker-optimal equilibrium under a production function of the form $f(h)$ is precisely the implicit tuition aq_k because this production function does not provide programs with infra-marginal productive rents. Under this production function, markdowns from output are determined only by residents' preferences for programs. Consequently, calculating firm profits using a production function of this type may provide a conservative approach to estimating payoffs to programs more generally. The next result shows that under heterogeneous preferences for programs, the difference between salaries and output is the same for all production functions of the form $f(h)$. This ensures that an implicit tuition can be defined and calculated using only the residents' willingness to pay for programs, circumventing the need for estimating f.

For notational simplicity, I state the result for a one-to-one assignment model, and the general result for many-to-one setting is stated and proved in Appendix B.3.4.[39] With a slight abuse of notation, let the total surplus from the pair $(i,\ j)$ be $a_{ij}^{f} = u_{ij} + f(h_i) \geq 0$.[40] Here, u_{ij} is the utility, net of wages, that resident i receives from matching with program j and $f(h_i)$ is the output produced by resident i. I now characterize the equilibria for a modified assignment game in which the surplus produced by the pair is $a_{ij}^{\tilde{f}} = u_{ij} + \tilde{f}(h_i) \geq 0$ in terms of the equilibria of the game with surplus a_{ij}^{f}.

Proposition 2.2. *The equilibrium assignments of the games defined by a_{ij}^{f} and $a_{ij}^{\tilde{f}}$ coincide. Further, if u_i^f and v_j^f are equilibrium payoffs for the surplus a_{ij}^{f}, then $u_i^{\tilde{f}} = u_i^f + \tilde{f}(h_i) - f(h_i)$ and $v_j^{\tilde{f}} = v_j^f$ are equilibrium payoffs under the surplus a_{ij}^{f}. Hence, a firm's profit in a*

[39]In the general formulation, I assume that the total output from a team of residents $(h_1, \ldots, h_{q_j})$ is $F(h_1, \ldots, h_{q_j}) = \sum_{k=1}^{q_j} f(h_k)$, where $f(h_k) = 0$ if position k is not filled.

[40]This formulation implicitly assumes that, at every program, it is individually rational for a worker to accept a salary equal to her product. It further assumes that the output of every resident is non-negative.

worker-optimal equilibrium depends on $\{u_{ij}\}_{i,j}$ *but is identical for all production functions of the form* $f(h)$.

Proof. See Appendix B.3.4 for the general case with many-to-one matches. □

As in the illustrative model, under a production technology that depends only on human capital, the residents are the residual claimants of output. An increase or decrease in the productivity of human capital is reflected in the wages, one for one. The firms' profits depends only on the preferences of the residents. Thus, I refer to the difference between output and salaries in the worker-optimal competitive equilibrium for a model in which f depends only on h as the implicit tuition. This definition uses the assumption that preferences of the programs can be represented using a single human capital index in the empirical model but also makes the additional restriction that the productivity of human capital, in dollar terms, does not depend on the identity of the program.

To the best of my knowledge, a closed form expression for competitive equilibrium salaries is not available when preferences of the residents are heterogeneous. I calculate the implicit tuition implied by estimated preferences using a two-step procedure.[41] Each step solves a linear program based on the approach developed in Shapley and Shubik (1971):

- *Step* 1 : Solve the optimal assignment problem, modified from the formulation by Shapley and Shubik (1971) to allow for many-to-one matching.
- *Step* 2 : Calculate the worker-optimal element in the core given the assignments from step 1.

Appendix B.3.3 describes the procedure in more detail. All calculations are done with the 2010-2011 sample of the data.

Table 2.10: Implicit Tuition

	Full Heterogeneity (1)	Geographic Heterogeneity (2)	Geo. Het. w/ Wage Instrument (3)
Mean	$23,802.64	$22,627.64	$43,470.39
	(5526.15)	(3495.62)	(13678.08)
Median	$21,263.30	$21,167.71	$40,606.85
	(5076.79)	(3265.54)	(12847.51)
Standard Deviation	$16,661.17	$12,278.42	$24,792.30
	(3946.33)	(1781.09)	(7485.20)
5th Percentile	$2,795.23	$5,179.08	$7,912.03
	(1008.51)	(1441.71)	(3246.19)
25th Percentile	$11,648.70	$14,070.10	$24,853.10
	(2820.62)	(2364.41)	(8299.05)
75th Percentile	$31,467.42	$28,902.46	$58,354.66
	(7131.65)	(4347.95)	(18134.03)
95th Percentile	$55,279.76	$45,784.76	$92,343.91
	(12758.48)	(6921.96)	(28071.67)

Notes: Based on 100 simulation draws. Each simulation draws a parameter from a normal with mean $\hat{\theta}_{MSM}$ and variance $\hat{\Sigma}$, where $\hat{\Sigma}$ is estimated as described in Section 2.6.4. Standard errors in parenthesis.

2.8.3 Estimates of Implicit Tuition

Estimates presented in Section 2.7 suggest that residents are willing to take large salary cuts in order to train at more preferred programs, which can translate into a large implicit tuitions. Table 2.10 presents summary statistics of the distribution of implicit tuition using estimates from specifications (1) through (3). I estimate the average implicit tuition to be about \$23,000 for specifications (1) and (2). This estimate rises to \$43,500 when using the instrument in specification (3) because the coefficient on salaries falls. As mentioned in Section 2.7, the instrument used appears weak and yields non-robust point estimates, but generally results in a larger willingness to pay and implicit tuitions through a decrease in the coefficient on salaries.[42] The standard error in the estimate using specification (3) is also large, at \$13,700, but can rule out an average implicit tuition smaller than \$17,000. These estimates are economically large in comparison to the mean salary of about \$47,000 paid to residents.

The results also show significant dispersion in the implicit tuition across residents and programs. The standard deviation in the implicit tuition is between \$12,000 and \$25,000. The 75th percentile of implicit tuition can be about three times higher than the 25th percentile, with even higher values at the 95th percentile. This dispersion primarily arises from the differences in program quality, which allows higher quality programs to lower salaries more than relatively lower quality program.

The estimated implicit tuition is between 50% to 100% of the \$40,000 salary difference between medical residents and physician assistants. This finding refutes the plaintiffs' argument that the salary gap would not exist if residents' salaries were set competitively and physician assistant salaries approximated the productivity of residents. However, the estimated implicit tuition cannot explain the salary gap between starting physicians and

---

[41] Since the total number of residents observed in the market is less than the number of positions and the value of options outside the residency market are difficult to determine, I will assume that the equilibrium is characterized by full employment. This property follows if, for instance, it is individually rational for all residents to be matched with their least desirable program at a wage that is equal to the total product produced by the resident at this program and the product produced by a resident is not negative.

[42] The instrumented version of specification (1) results in implicit tuition estimates much larger than the ones reported because of the smaller estimated coefficient on salaries.

medical residents, which is approximately $90,000.[43] As discussed earlier, the implicit tuition is a conservative estimate of the salary markdown and part of this salary gap may be due to differences in the productivity of medical residents and starting physicians.

When residents' preferences are heterogeneous, the implicit tuition is also a function of the relative demand and supply of different types of residency positions, and is not simply a result of compensating differentials. Estimates from specification (1) imply a willingness to pay by residents for programs in the same state as their medical school, and programs in the same state as their birth state. Therefore, the demand for residency positions is high in states where many residents were born or states where many residents went to medical school. A supply-demand imbalance occurs, for instance, when the number of residency positions in the state is low but many residents have preference for training in that state. These forces will be important determinants of equilibrium salary if the residency market adopts the design proposed in Crawford (2008) because the proposal is intended to produce a competitive equilibrium outcome.

To demonstrate the effect of this imbalance on the estimated implicit tuition, I present results from the regression

$$\ln y_j = z_j \rho_1 + \rho_2 \ln npos_{s_j} + \rho_3 \ln gr_{s_j} + \rho_4 \ln born_{s_j} + e_j,$$

where y_j is the average implicit tuition at program j estimated using specification (1), z_j are characteristics of program j included in specification (1), s_j is program j's state, $npos_{s_j}$ is the number of residency positions offered in s_j, gr_{s_j} is the number of residents from MD medical schools in state s_j and $born_{s_j}$ is the number of residents born in state s_j. Column (4) of Table 2.11 shows that the elasticity of the average implicit tuition at a program with respect to the number of family medicine graduates getting their degrees in a medical school in that state is positive, $\hat{\rho}_3 = 0.19$. Conversely, the elasticity with respect to the number of positions offered in the program's state is negative, $\hat{\rho}_2 = -0.16$. The estimate for $\hat{\rho}_4$ is not statistically significant, partially because the estimated preference for birth state is low and because supply-demand imbalance based on birth-state is also lower.

[43] I use Mincer equation estimated using interval regressions on confidential data from the Health Physician Tracking Survey of 2008 to calculate the average salaries for starting family physicians. Details in Appendix B.6.

Table 2.11: Dependence of Implicit Tuition on Demand-Supply Imbalance

	Log Average Implicit Tuition in Program			
	Full Heterogeneity			
	(1)	(2)	(3)	(4)
Log Residency Positions	0.0008	-0.1557***	-0.0578***	-0.1442***
in Program State	(0.0044)	(0.0106)	(0.0101)	(0.0128)
Log Family Medicine MD Graduates		0.1851***		0.1951***
from Program State		(0.0114)		(0.0130)
Log US Born Residents			0.0658***	-0.0233
in Program State			(0.0102)	(0.0145)
R-squared	0.4144	0.4180	0.4150	0.4180

Notes: Linear Regressions. Dependent variable is the log of total implicit tuition at a residency program divided by the number of residents matched to the program. All regressions on generated implicit tuitions data using the 2010-2011 sample of residents and programs, and 100 simulation draws. All regressions include Log Beds, Log NIH Fund (Major), Log NIH Fund (Minor), dummies for no NIH funded affiliated, Medicare Case Mix Index, Rural Program dummy and Program type dummies. Standard errors clustered at the simulation level. Significance at 90% (*), 95% (**) and 99% (***) confidence.

2.8.4 Discussion

In matching markets, agents on both sides are heterogeneous and have preferences for match partners. The effects of this feature on market outcomes, especially when barriers to entry are substantial, are not captured by a perfect competition model. Theoretical results presented in Section 2.8.1 show that equilibrium salaries can be well below the product of labor, net of costs of training, when residents value the quality of a program. Counterfactual estimates show that the willingness to pay for programs results in a large markdowns in salaries in a competitive wage equilibrium. The upper end of estimates can explain the salary gap between physician assistants and medical residents assuming that physician assistant salaries are close to the productivity of residents. My estimates also show that higher quality programs would earn a larger implicit tuition than less desirable programs. To the extent that higher quality programs are matched with higher skilled residents and are also instrinsically more productive, the implicit tuition is a countervailing force to high dispersion salaries driven by productivity differences.

The analysis suggests that instead of the design of the match, salaries are low because programs are capacity constrained and barriers to entry are large due to fixed costs or accreditation requirements. The implicit tuition can therefore explain the empirical observations of Niederle and Roth (2003a, 2009) in fellowship markets and highlights why analyzing matching markets using a perfect competition model can be quantitatively misleading.

In this market, salaries may also be influenced by the previously mentioned guideline requiring minimum financial compensations for residents. While these forces may be important, they seem unrelated to the match. In other words, programs may not have the incentive to pay salaries close to levels suggested by the plaintiffs because of economic primitives.

2.9 Application 2: Rural Hospitals

Access to medical care is significantly lower in rural communities of the United States: about a fifth of the US resides in rural counties but only a tenth of physicians practice in these areas(Rosenblatt and Hart, 2000). Increasing residency training in rural areas is seen as an important part of solutions to this disparity in access to care because of the empirical association between rural training or background with recruitment and retention of rural physicians (Brooks et al., 2002; Talley, 1990). About 20% of urban born residents graduating from family medicine programs start their initial practice in rural areas, roughly in proportion to the population in rural communities of the US, whereas about 46% of rural born family medicine residents begin their practice in rural communities (Table B.5). Both urban-born and rural-born residents trained in rural areas are about 30 percentage points more likely to enter a rural practice after residency (Table B.5). While some of this association is probably driven by selection into rural residency training programs, it may also partly be a causal effect of rural training. The difference in the nature of urban and rural medicine and specialized experience useful for practicing in rural areas may be a contributing factor.[44]

The Patient Protection and Affordable Care Act of 2010 (ACA) contains provisions for increasing the training and recruitment of primary care physicians in rural areas. The ACA provides an additional $1.5 billion to loan forgiveness programs focussed on recruiting physicians into health physician shortage areas and creates targeted grants for increasing residency training positions in primary care, especially in rural areas.[45] Similar concerns motivated Japan to institute regional caps that reduced the number of positions in urban programs proportionally to their size. Arguably, caps on urban programs could be implemented in the United States through the Accreditation Council for Graduate Medical Education (ACGME). In fact, the ACA moves a large number of unused Medicare funds allocated for supporting costs of residency training in urban programs to states with disproportionately low resident-to-population ratios and rural areas (see §5503 ACA, 2010).

---

[44]Non-specialist primary care physicians tend to supply a disproportionately larger fraction of medical care in rural counties, including emergency and obstetrics care. Family medicine residents training in rural areas may consequently be more likely to receive specific experience for practicing rural medicine. Many practitioners concerned with the rural physician shortage argue for an increased emphasis on rural residency training through either rural programs or rotations (Rabinowitz et al., 2008; Rosenblatt and Hart, 2000).

[45]The ACA supplements the budget of the National Health Services Corps loan forgiveness program. Section 5301 provides grants for enhancing capacity at existing primary care training locations and Sections 10501 (I) 5508(a) provides grants specifically for establishing new programs in rural health clinics and programs. See Bailey (2010) or Table 2 of the Congressional Research Service report titled ”Discretionary Spending in the Patient Protection and Affordable Care Act (ACA).”

Broadly speaking, the ACA enacts recruitment incentives and quantity regulations to encourage physician supply in rural areas. I study the effects of these policies by comparing simulated outcomes from environments with and without the intervention. A complete model of the market makes it possible to account for general equilibrium effects. I focus on quantifying impact of these policy interventions on the sorting and number of residents in rural programs because many of the private and social costs and benefits are difficult to quantify. Insight on the assignments resulting from these interventions may influence the decisions of a social planner considering such policies.

All simulations are conducted using the 2010-2011 academic year of the data and specification (1). I assume that the policies do not affect the entry of residents into the market. Specifications (2) and (3) yield qualitatively similar results. Specification (1) does not use an instrument for salaries, which Section 2.7 notes is likely to result in an overestimate of the coefficient on salaries. This is not a primary concern in the analysis of supply interventions because salaries are kept fixed, and only the choices residents conditional on salaries are important. The analysis of financial incentives, however, may overestimate the sensitivity of residents to these policies.

2.9.1 Financial Incentives for Rural Training

I mimic the loan forgiveness programs of the National Health Services Corps, except for medical residents. The program currently provides an annual incentive of \$20,000 to \$30,000 to primary care physicians for practicing in Health Physician Shortage Area, usually rural or inner-city communities. To simulate the impact of such recruitment incentives for residents training in rural areas, I exogenously increase the salaries at rural hospitals by \$5,000, \$10,000 and \$20,000. The average estimated utility difference between the rural and urban programs is between \$5,000 and \$10,000 (Table 2.9).

Panel A of Table 2.12 presents summaries from the baseline simulation from the model using data from the year 2010 - 2011. The number of positions filled in rural areas, as observed in the data, is 310. The average predicted by the model is slightly higher at 313.37 although the inter-quartile range of simulations contains the observed number of matches. According to baseline simulations, the quality of doctors matched with rural areas is similar

Table 2.12: Effects of Policy Instruments for Encouraging Rural Training

	Full Heterogeneity (Specification 1)		
Panel A: Baseline Simulations (310/334 positions filled in data)			
Simulated Matches	313.33		
	(310 - 317)		
Prob. Rural Match > Urban Match	52.76%		
Panel B: Salary Incentives	\$5,000	\$10,000	\$20,000
	(1)	(2)	(3)
Rural Matches	10.23	17.3	20.63
	(7 - 12)	(14 - 21)	(17 - 24)
Δ Prob. Rural Match > Urban Match	9.38%	17.70%	31.28%
Total Cost of Subsidy (mil.)	\$1.62	\$3.31	\$6.68
Δ Private Welfare of Residents (mil.)	\$1.84	\$3.64	\$7.05
Cost Per Additional Resident	\$158,143	\$191,116	\$323,762
Panel C: Quantity Regulations	Decrease urban proportionally (i)	+2 positions for rural programs (ii)	Combination of (i) and (ii) (iii)
Modified Urban Capacity	2846	2963	2688
%Δ in Urban Capacity	-3.95%	—	-9.28%
Modified Rural Capacity	334	460	460
%Δ in Rural Capacity	—	37.72%	37.72%
Δ in # Rural Matches	12.01	121.31	146.63
	(4.5 - 20)	(114.5 - 128)	(137.5 - 156.5)
Δ Prob Rural Match > Urban Match	-0.56%	7.02%	-3.73%
Δ Residents' Private Welfare (mi)	-\$3.76	\$5.39	-\$5.49

Notes: In Panel C, Column (i) decreases the urban positions in proportion to program size, subject to integer constraints. Positions at urban programs were reduced in proportion until further reductions would yield a greater number of residents than programs. In column (i), this yielded 32 more positions than residents. In column (iii), the number of residents equals the number of positions. All simulations use 2010 - 2011 sample with 3,148 residents and 3,297 total number of positions. Baseline and counterfactual simulations using 100 draws of structural unobservables. Inter-quartile range in parenthesis. Prob. X > Y is the Wilcoxian statistic: probability that the human capital population X is drawn from is greater than that of the population that Y is drawn from.

to the quality of doctors in urban areas. This is consistent with the reduced form evidence presented in Table 2.3 that do not see a significant disadvantage to currently operating rural programs.[46]

Panel B presents the impact of increased incentives for rural training. The incentive affects residents roughly indifferent between a rural and an urban program to rank the rural program ahead of the urban program. Across the board, we see small increases in the number of residents matches to programs in rural communities. An incentive of \$20,000 increases the number of residents training in rural areas by about 17, or 5.5% of the number of positions in rural programs. This incentive costs the government \$325,000 per additional resident matched to a rural program because most of the loan forgiveness accrues to residents assigned to positions that would be occupied without the financial incentive. Instead of affecting numbers, the primary impact of incentives is an increase in the human capital of residents matching to rural areas. As compared to a baseline of about an even chance, under a small \$5,000 incentive, a randomly chosen rural resident is about 9.4 percentage points more likely to have a higher human capital than an urban resident. This increase in the quality of residents is increasing with size of the incentives.

These results can be explained by capacity constraints in rural areas. While price incentives directly increase the number of residents ranking rural programs ahead of urban programs, the number that match with any given program is constrained by its capacity. With 310 out of 334 positions filled, there is little scope for the incentive to substantially increase numbers. Consequently, although the incentives increase the pool of residents ranking rural programs higher, capacity constraints prevent an increase in numbers but allow an increase in the quality of residents matched at subsidized programs.

One may ask whether a simpler analysis based on partial equilibrium reasoning with unilateral salary increases by programs would lead to similar conclusions on the assignments between residents and programs. The quasi-linear utility function implies that a uniform increase in salaries of all residency programs would not impact assignments because the comparison between any two programs remains unchanged. A partial equilibrium

[46]Unconditionally, rural programs are 7 percentage points more likely to be matched with residents that have an MD degree. The average medical school median MCAT score of a resident matched with a rural programs is less than a point lower, and the average NIH funding is 0.3 log points lower.

analysis based on unilateral salary increases substantially deviates from this prediction. For smaller interventions we expect general equilibrium effects to be less pronounced. In Appendix B.4.2, I compare general and partial equilibrium effects of incentivizing rural training, and more broadly, training in medically underserved states. I find that a partial equilibrium analysis overestimates the number of positions allocated for small incentives, but for larger incentives, overestimates aggregate increases in the quality of residents.

Welfare Effects and the Importance of Heterogeneity

It is not obvious whether the small increase in numbers and a larger increase in the quality of residents matched with rural programs is socially desirable. A complete cost-benefit analysis depends on the private surplus to programs and residents as well any social benefits of rural training. The model only allows us to quantify the cost of financial incentives and its impact on the total private surplus to residents. Table 2.12 shows that a \$5,000 incentive results in a transfer of \$1.6 million from the government to residents. However, the estimated increase in residents' private welfare is 13.5% more than this amount. This result is a consequence of heterogeneous preferences and the ability of financial incentives to realize potential efficiency gains by assigning residents with the lowest distaste for rural programs to those positions. A small incentive for training in a rural program only induces a resident who is roughly indifferent between a rural and an urban program to choose rural training. This resident then opens up a position in an urban program that may be strongly preferred by another resident. Therefore, general equilibrium re-sorting effects of the financial incentive result in an increase the efficiency of assignments.

The potential for financial incentives for targeting residents with low distaste for rural areas only exists when preferences are heterogeneous. In a model that does not allow for heterogeneity, the willingness to pay for training at a program is identical across residents. Such a model would predict that a permutation of the assignment does not affect residents' welfare. The impact on the private benefits to residents, net of the transfer, is only through the total number of positions filled at different programs.

2.9.2 Supply Interventions

I assess the impact of supply regulations in this market by simulating outcomes after changing the number of positions offered at different programs. I consider three types of policy interventions. The first mimics the policy implemented in Japan and reduces the number of positions in urban programs proportional to the size of the program (subject to integer constraints) until further reductions would lead to fewer positions than the total number of residents in the market. The second intervention is motivated by the provisions in the ACA for increasing the number of rural training positions. Since the characteristics of new programs are not known, I increase the number of positions in existing rural programs. This can be thought of as creating copies of existing programs via grants funded by the ACA. The final intervention combines the two by first increasing the number of positions at existing rural programs followed by decreasing the number of positions in urban programs proportionally. In all counterfactuals, the number of residents and observed characteristics are the same as in the dataset. Consequently, the second intervention has significantly more positions than programs.

Panel C of Table 2.12 presents the estimated effects of these policy interventions. Since a policy that reduces the number of positions offered at urban programs displaces residents from urban areas, it mechanically increases the number of residents matching at rural programs. However, the sorting effects of these changes are not *a priori* clear. A naive reasoning may lead to the conclusion that caps have a large adverse impact on the quality of residents training at rural programs because displaced residents are disproportionately less desired by the programs they are matched to. However, residents displaced from urban programs in turn displace others, resulting in overall resorting. According to estimates from both models, the distribution of resident quality matching at rural programs is similar to the distribution before the caps.

A major, perhaps not surprising, impact of the caps is the loss in private welfare of residents from the decreased availability of positions. This decrease results in a similar number of additional residents in rural programs as a $5,000 financial incentive. However, price incentives result in an overall gain for residents in addition to the transfer. The observation suggests that quantity regulations are a blunt policy instrument that do not target residents with the least dislike for rural positions.

Column (ii) presents the impact of increasing the number of positions in rural residency programs by two each. This policy significantly increases the number of residents matched to rural programs and also results in an increase in the quality of residents in rural areas. As compared to outcomes prior to the policy, the typical residents assigned to a rural program is 7 percentage points more likely to have a higher human capital index than a resident matched to an urban program. The change in quality of residents in rural areas is due to increases in the number of residents matched at the highest quality rural programs but decreases in the number of residents matched at low quality residency programs in urban and rural areas. Although not considered here, entry of additional residents into the family residency market could mitigate adverse effects of unfilled positions.

Finally, the third policy combines the other two and, by construction, has a large effect on the number of residents placed in rural programs. As compared to a singular increase in positions offered in rural areas, this policy can adversely affect the quality of residents assigned to rural programs. The reason is that residents with a low human capital are forced into undesirable residency positions that were earlier left vacant under an increase in rural positions.

2.9.3 Discussion

Many regulations target an activity in which levels alone determine social benefits. In the context of residency training and other matching markets, a social planner may be concerned about the type of resident training in a rural area in addition to the total number of residents. For instance, if retention is an important goal, we may prefer a policy that yields residents with higher intrinsic preference for rural areas in rural training locations. The costs imposed on urban programs by these interventions are yet another factor that may influence optimal policy design. The analysis presented here sheds light on general equilibrium sorting impacts of interventions that should be considered when designing policy towards rural training.

The exercise also illustrates the ability of the model to understand policy interventions in matching markets more broadly. In settings where sorting may be an important consideration in policy decisions, the methods developed in this paper are a natural tool for analysis. There are perhaps other equally important factors influencing policy choices, such as the endogenous decisions of participating in the market or setting salaries. It may be possible to use an appropriately augmented version of this model to incorporate such decisions. In this study, I hold these decisions held fixed to narrowly focus on the direct effects of studied interventions.

2.10 Conclusion

Two key features of two-sided matching markets are that agents are heterogeneous and that highly individualized prices are often not used. Both properties have important implications for equilibrium outcomes because assignments are determined by the mutual choices of agents rather than price-based market clearing. A quantitative analysis of policy interventions may therefore require estimates of preferences on both sides of the market.

When data on stated preferences is available, extensions of discrete choice methods can provide straightforward techniques for analysis (see Abdulkadiroglu et al., 2012; Hastings et al., 2009, among others). A common constraint is that only data on employer-employee matches or student enrollment records, rather than stated preferences, are available. This paper develops empirical methods for recovering preferences of agents in two-sided markets with low frictions using only data on final matches. I use pairwise stability together with a vertical preference restriction on one side of the market to estimate preference parameters using the method of simulated moments. The empirical strategy is based on using sorting patterns observed in the data and information available only in many-to-one matching. Sorting patterns alone cannot be used identify the parameters of even a highly simplified model with homogeneous preferences on both sides of the market.

These methods allow me to empirically analyze two important issues concerning the market for medical residents. First, I address the academic debate on whether centralization in this market causes low salaries. A stylized model shows that a limited supply of desirable residency positions can depress salaries even under frictionless competitive negotiations. Residents' willingness to pay for desirable programs results in average salaries that are at least \$23,000 lower than levels suggested by a perfect competition model. Models using wage instruments result in imprecise but higher estimated markdowns, of about \$43,000. These markdowns are due to an implicit tuition that can explain the gap between incomes of medical residents and physician assistants, and also the empirical observations of Niederle and Roth (2003a, 2009). The result suggests that the limited supply of heterogeneous residency positions is the primary cause of low salaries, and weighs against the view the match is responsible for low resident salaries.

Second, I show that policy interventions aimed at encouraging rural training have important effects on the sorting of residents. For this reason, price incentives and quantity regulations are not equivalent policy instruments. Furthermore, the size, scope and design of these interventions significantly influence the qualitative and quantitative effects of these interventions. While supply regulations are more effective at increasing the number of residents in rural areas, financial incentives are able to specifically target residents that do not significantly dislike training in rural areas. Analyzing the general equilibrium effects of both interventions on residents' private welfare and the sorting of residents into rural areas needs a complete model of market primitives.

The methods and analysis in this paper can be extended in several directions. The restriction on the preferences of one side of the market could be relaxed in other markets if the data contain information that would allow estimating heterogeneous preferences on both sides of the market. For instance, it may have been possible to estimate heterogenous preferences for residents if program characteristics that can plausibly be excluded from resident preferences were observed. Future research in other matching markets could use data from several markets in which the composition of market participants differs in order to estimate heterogeneous preferences on both sides. These extensions must also confront methodological hurdles arising from a multiplicity of equilibria are important in other matching markets.

General equilibrium effects of price and supply interventions are important in other matching markets as well. For instance, tuition regulations in public universities and public school reforms introducing new schools or shutting down under-performing schools also affect the sorting of students. There are also additional effects of these policies on other endogenous choices such as entry decisions and price or capacity setting. In future research, I plan to use theoretical and empirical tools to further investigate these interventions in matching markets.

Chapter 3

Identification and Estimation in Two-Sided Matching Markets

3.1 Introduction

There has been recent interest in estimating preferences in two-sided matching markets for use in counterfactual policy analysis.[1] These models typically use data from observed match outcomes and an equilibrium assumption on these matches in order to infer preferences of agents in the market. However, most formal econometric results have been restricted to a model with transferable utility, in which only the aggregate surplus generated by the agents is identified (See Choo and Siow, 2006, for example).

This paper presents the first theoretical results on identification and estimation in a matching market with non-transferable utility. We use pairwise stability as an equilibrium concept to model the observed outcomes and infer preferences (Roth and Sotomayor, 1992). Under this concept, no two unmatched pair of agents have the incentive to match with each other over their assigned partners. Our model considers the restrictive case in which preferences on each side of the market are vertical, i.e. all agents agree on the relative ranking of any two agents.

We study a data environment in which matches from a single large market is observed.

[1] (See Fox, 2009, for a survey)

A main result is that the distribution of preferences on both sides of the market are not identified when data from one-to-one matches are observed. In contrast, we show that data from many-to-one matches can be used to non-parametrically identify the distribution of preferences on both sides of the market. Further, the additional identifying information available in many-to-one matches can be clearly illustrated using simulations. We find than an objective function that only uses information available in sorting patterns is not able to distinguish between a large set of parameter values. In data from one-to-one matching, this is the only information known in the dataset. In contrast, we show that an objective function that uses information in many-to-one matching does not suffer from this problem and has a global minimum near the true parameter.

We then study asymptotic properties of a method of moments estimator based on a criterion function that uses moments from many-to-one matching as well as sorting patterns. The main result proves that under identifiable uniqueness of a parametric family of models, data on a single large market with many-to-one matches can be used to consistently estimate the true parameter. For simplicity, we restrict to the case with two-to-one matching. Finally, we use Monte Carlo simulations to study the property of a simulation based estimator.

Most of the previous literature studies the case in which the goal is to recover a single aggregate surplus that is split between agents (Chiappori et al., 2011; Choo and Siow, 2006; Fox, 2008; Galichon and Salanie, 2010; Gordon and Knight, 2009; Sorensen, 2007). A constraint faced by these studies is that monetary transfers between matched partners are not observed, providing limited hope for estimating two separate utility functions is limited. In some applications, such as the matching of medical residents studied in Agarwal (2012), monetary transfers are observed and counterfactual analysis requires estimates of both distribution of preferences.

Our work is related to Logan et al. (2008) and Menzel (2011), which propose Bayesian techniques for estimating the posterior distribution in marriage markets using sorting patterns observed in the data. Boyd et al. (2003) estimate the preferences of teachers for schools and schools for teachers. These papers use sorting patterns observed in the data to recover primitives because standard revealed preference arguments do not apply. Our result on non-identification with one-to-one matching questions these approaches and recommend that information in many-to-one matches be used for estimation.

Section 3.2 presents the model, Section 3.3 discusses identification, Section 3.4 discusses estimation results and Section 3.5 presents Monte Carlo results. All proofs are in the Appendix.

3.2 Model

We will consider a two-sided matching market with non-transferable utility. The two sides will be referred to as workers and firms indexed by i and j respectively. For simplicity, we assume that the total number of positions at firms equals the total number of workers. Additional firms could be introduced in order to capture unmatched workers.

3.2.1 Market Participants

The participants in the market are defined by a pair of probability measures $m^e = (m^{x,\varepsilon}, m^{z,\eta})$. Here, $m^{x,\varepsilon}$ is the joint distribution of observable traits $x \in \chi \subseteq \mathbb{R}^{k_x}$ and unobsevable traits $\varepsilon \in \mathbb{R}$ for the workers. Likewise, $m^{z,\eta}$ is the joint distribution of observable traits $z \in \zeta \subseteq \mathbb{R}^{k_z} \times \mathbb{N}$ and unobservable traits $\eta \in \mathbb{R}$ for the firms.

In an economy with n agents $((X_1, \varepsilon_1), \ldots, (X_n, \varepsilon_n))$ and $((Z_1, \eta_1), \ldots, (Z_n, \eta_n))$ on each side, the measures will be of the form $m_n^{x,\varepsilon} = \frac{1}{n} \sum_{i=1}^{n} \delta_{(X_i, \varepsilon_i)}$ and $m_n^{z,\eta} = \frac{1}{n} \sum_{j=1}^{n} \delta_{(Z_j, \eta_j)}$ where δ_Y is the dirac delta measure at Y.

Assumption 3.1. *The population measures $m^{x,\varepsilon}$ and $m^{z,\eta}$ satisfy*

(i) m^η and m^ε admit densities will full support on $\mathbb{R}$, and are absolutely continuous with respect to Lebesgue measure.

(ii) $m^{\eta|x} = m^\eta$ for all $x \in \chi$ and $m^{\varepsilon|z} = m^\varepsilon$ for all $z \in \zeta$

Assumption 3.1 (i) imposes a regularity condition on the support and distributions of the unobservables and Assumption 3.1 (ii) assumes independence. On its own, independence is not particularly strong, but a restriction on preferences to follow will make this a strong assumption.

3.2.2 Preferences

Each side of the economy has a utility function over observable and unobservable traits of the other side of the economy. That is, worker i's human capital index is given by the additively separable form

$$v(x_i, \varepsilon_i) = h(x_i) + \varepsilon_i. \tag{3.1}$$

Likewise, the preference of workers for firm j is given by

$$u(z_j, \eta_j) = g(z_j) + \eta_j \tag{3.2}$$

where $g(\cdot)$ does not depend on the number of seats.

In addition to homogeneity, additivity of $v(x_i, \varepsilon_i)$ in ε_i and of $u(z_j, \eta_j)$ in η_j are strong assumptions when ε_i and η_j are independently distributed of x_i and z_j. While this is difficult to economically motivate, it is commonly used in discrete choice literature. This paper is a first step towards providing theoretical results on identification and estimation in this market, and these assumptions significantly ease the analysis.

3.2.3 Pairwise Stability

Definition 3.1. *A **match** is a probabiliy measure μ on $(\chi \times \mathbb{R}) \times (\zeta \times \mathbb{R})$ with marginals $m^{x,\varepsilon}$ and $m^{z,\eta}$ respectively.*

The traditional definition of a match used in Roth and Sotomayor (1992) is based on a matching function $\mu^*(i) \mapsto J \cup \{i\}$ where J is th set of firms. With probability 1, such a function defines a unique counting measures of the form $\mu_n = \frac{1}{n}\sum_{i,j=1}^{n} \delta_{(X_i, \varepsilon_i, Z_j, \eta_j)}$ where $\delta_{(X_i, \varepsilon_i, Z_j, \eta_j)} > 0$ only if i is matched to j in a finite sample. This fact is a consquence of Assumption 3.1 (i), which implies that in a finite economy, (z, η) identifies a unique firm with probability 1.[2]

Definition 3.2. *A match μ is **pairwise stable** if there do not exist two (measurable) sets $S_I \subseteq \chi \times \mathbb{R}$ and $S_J \subseteq \zeta \times \mathbb{R}$ in the supports of $m^{x,\varepsilon}$ and $m^{z,\eta}$ respectively, such that $\int_{S_I} v(X, \varepsilon)\, dm^{x,\varepsilon} > \int_{S_I} v(X, \varepsilon)\, d\mu(\cdot, S_J)$ and $\int_{S_J} u(Z, \eta)\, dm^{z,\eta} > \int_{S_J} u(Z, \eta)\, d\mu(S_I, \cdot)$.*

[2]In addition to a traditional matching function, in finite sample our definition also allows for fractional matchings. However, such realizations are not observed in realized datasets on matches.

This definition of pairwise stability is also equivalent to that for a finite market since μ may not have support on blocking pairs. Existence of a pairwise stable match follows in a finite market because preferences are responsive (Roth and Sotomayor, 1992) and uniqueness follows from alignment of preferences as disscussed in Clark (2006) and Niederle and Yariv (2009).

Remark 3.1. *In the model employed here the pairwise stable match μ has support on $(x, \varepsilon, z, \eta)$ only if $F_U(u(z, \eta)) = F_V(v(x, \varepsilon))$ where F_U and F_V are the cumulative distributions of u and v respectively.*

3.3 Identification

This section presents conditions under which we can the identify the functions $h(x)$, $g(z)$ and the distributions of ε and η using the marginal distribution of the match μ on the observables, $\chi \times \zeta$. These objects allow determine the distribution of preferences, or the probabilities

$$\mathbb{P}(v(X_1, \varepsilon_1) > v(X_2, \varepsilon_2) | X_1 = x, X_2 = x') \text{ and } \mathbb{P}(u(Z_1, \eta_1) > u(Z_2, \eta_2) | Z_1 = z, Z_2 = z'). \tag{3.3}$$

We make the following assumptions on $h(\cdot)$ and $g(\cdot)$

Assumption 3.2. *(i) $h(\bar{x}) = 0$, $|\nabla h(\bar{x})| = 1$ and $g(\bar{z}) = 0$, $|\nabla g(\bar{z})| = 1$, and η and ε are median zero.*

(ii) ε and η are median-zero

(iii) $h(x)$ and $g(z)$ have full support over $\mathbb{R}$

(iv) $h(\cdot)$ and $g(\cdot)$ are differentiable

(v) The measures m^x and m^z admit densities f_X and f_Z

(vi) The densities f_ε and f_η are differentiable and have non-vanishing characteristic functions

Assumptions 3.2 (i) and (ii) impose scale and location normalizations that are necessary since the latent variables are not observed. Such normalizations are necessary in single-agent discrete choice models and are without loss of generality. Assumption 3.2 (ii) is a support condition often necessary for non-paramteric identification. Assumption 3.2 (iv) - (vi) are regularity conditions.

Assumption 3.3. *At the pairwise stable match μ, the conditional distribution $\mu^{(z,\eta)|(x,\varepsilon)} = \mu^{(z,\eta)|(x',\varepsilon')}$ if $h(x) + \varepsilon = h(x') + \varepsilon$ and $\mu^{(x,\varepsilon)|z,\eta} = \mu^{(x,\varepsilon)|(z',\eta')}$ if $g(z) + \eta = g(z') + \eta'$.*

This assumption requires that the desirability of an agent in the market is a sufficient statistic for their matches. In other words, the sorting observed in the data can depend only on the observable characteristics through their effect on the desirability to the other side of the market. Without this assumption, sorting patterns in the data may not be related to preferences whatsoever.

3.3.1 Identification from Sorting Patterns

Our first result shows that $h(\cdot)$ and $g(\cdot)$ are identified up to monotone transformations using only sorting patterns in the data.

Lemma 3.1. *Under Assumptions 3.1 and 3.3, and the representation of preferences in equations (3.1) and (3.2), the level sets of the functions $h(\cdot)$ and $g(\cdot)$ are identified a one-to-one observed match μ.*

Proof. See Appendix C.1.1. □

We can detemine whether or not two worker types x and x' are equally desirable from the sorting patterns observed in one-to-one, hence also in many-to-one matches. Intutively, if two worker types have equal values of $h(\cdot)$, then the distributions of the desirability of firms they match with are identical. In a pairwise stable match, under the additive structure of equations (3.1) and (3.2), this also implies that the distribution of firm observable types these workers are matched with is identical. Conversely, if two worker types are matched with different distribution of firm observables, they cannot be identical in observable quality because of Assumption 3.3.

The result does not identify $h(\cdot)$ and $g(\cdot)$ upto positive monotone transformations on either side of the market. In particular, it does not tell whether a any given worker trait is desirable or not. Intuitively, assortative matching between, say firm size and worker age, may result from either both traits being desirable or both traits being undesirable. The next result shows that under a **sign restriction** only on one side of the market, both $h(\cdot)$ and $g(\cdot)$ are identified up to positive monotone transformations.

Assumption 3.4. *The function $h(x)$ is strictly increasing in its first argument, x_1. Further, x_1 has full support in $\mathbb{R}$, and $\lim_{x_1 \to \infty} h(x) = \infty$, $\lim_{x_1 \to -\infty} h(x) = -\infty$.*

Proposition 3.3. *Assumption 3.4 and the conditions in Lemma 3.1 determine $h(\cdot)$ and $g(\cdot)$ up to positive monotone transformations.*

Proof. See Appendix C.1.2. □

The sign restriction allows us to order the level sets of h. Workers at higher level sets of h also receive a more desirable distribution of firms. We can then use this to order the level sets of $g(\cdot)$ as well. A symmetric result would hold under a sign restriction on g. Assumption 3.4, as stated, is fairly strong although it is conceivable that it may weakened using a proof technique that stitches together information from various components of h.

The next result shows a limitation of empirical content in data from one-to-one matches. For this proposition, a model defined by equations (3.1) and (3.2) satisfying Assumptions 3.3, 3.1 and 3.2 will be referred to as a matching model. The result shows that in a data environment with one-to-one matches, the matching model is observationally equivalent to modified model in which $\varepsilon \equiv 0$.

Proposition 3.4. *If F_U^{-1}, F_V, h^{-1}, g and the density of η are twice continuously differentiable, data from one-to-one matches can be rationalized in a matching model with $\varepsilon \equiv 0$.*

Proof. The proof proceeds by re-writing the matching model with $\varepsilon \equiv 0$ as a transformation model of Chiappori and Komunjer (2008), which they show is correctly specified. See Appendix C.1.3 for details. □

The result shows that despite imposing additional regularity conditions, data from one-to-one matches can be rationalized using a model in which only one set of unobservables are present. In these data, only the joint distribution of observable characteristics given by the match or sorting patterns are known. Logan et al. (2008) and Boyd et al. (2003)

employ empirical strategies that only use sorting patterns to estimate preferences. The result suggests that their estimates may be relying strongly on parametric assumptions.

As shown in the next section, data from many-to-one matching markets has additional information that is useful for identification. The dataset used by Boyd et al. (2003) contains this information, but their empirical strategy does not take advantage of it.

Further, the result is not pathalogical, as Section 3.3.3 presents simulations in which a simple parametric model can be used to illustrate this non-identification result. The problem in these simulations is allieviated when data from many-to-one matching markets is observed and used.

3.3.2 Identification from Many-to-One Matches

We now show that data from many-to-one matching markets can be used to identify the model. We consider a limit dataset in which there are a large number of firms, and each firm has a large number of workers. In such a dataset, with finitely many large firms, the distribution of observable worker types X is identified for each firm j. Thus, the data consist of a measure over worker distributions $F^{W}_{X|j}$.[3] In addition, the data consists of a joint measure F_{XZ} of worker and firm observable types. Our first result shows that the distribution of $U|Z$ is identified in such a dataset.

Proposition 3.5. *Under Assumptions 3.1 - 3.2, the function $g(z)$, the density f_η are identified from many-to-one matching.*

Proof. See appendix C.1.5. □

Intuitively, the quality of workers at a more desirable firm is better than a less desirable firm. Under 3.3, the distribution of the observable quality $h(x)$ of workers at a more desirable firm is also higher. This fact allows us to stochastically order the worker distributions $F^{W}_{X|j}$ and identify the τ-th quantile of within firm worker distributions. We can then identify the probability that a firm with characteristic z that have worker distributions that have dominated worker distributions, and consequently the quantile distribution of $U|Z$.

[3]In the large firms limit, the data consists of a measure over worker distributions at a firm. Formally, let $\mathcal{C}^{\chi}$ be the space of cdfs on χ. The data is a measure on $\mathcal{C}^{\chi} \times \zeta$.

The rest of the proof uses a deconvolution argument relying on the additive separability of η and support conditions to identify $g(z)$ and f_η.

Finally, we show that the primitives determining the distribution of $V|X$ are identified.

Proposition 3.6. *The function $h(x)$ and the distribution f_ε are identified if f_η, $g(z)$ are known and Assumptions 3.1 - 3.2 are satisfied.*

Proof. See appendix C.1.6. □

Given that $g(z)$ and the distribution of η are identified, the proof proceeds by identfying the distribution of firm observables that are matched with a firm quality index u using a deconvolution argument. Quantile-quantile matching implies that the object allows identifying the probability that a worker with characteristics x is preferable to a firm with characteristics x'. The rest of the proof uses this quantity to recover $h(x)$ and f_ε using additive separability and support conditions.

The results presented above identify all relevant primitives of the model.

Theorem 3.1. *The functions $g(z)$ and $h(x)$, and the densities f_η and f_ε are identified from data on many-to-one matching if Assumptions 3.1 - 3.4 are satisfied.*

Proof. Follows from Propositions 3.3, 3.5 and 3.6. □

3.3.3 Importance of Many-to-one Match Data: Simulation Evidence

The identification results presented in the previous section relied on observing data from many-to-one matching, and shows that the model is not identified using data from one-to-one matches. In this section, we present simulation evidence from a parametric version of the model to elaborate on the nature of non-identification and to illustrate the importance of using information from many-to-one matching in estimation.

We simulate a dataset of pairwise stable matches from a simple model and then compare objective functions of a method of simulated moments estimator that is constructed from moments only using information present in sorting patterns to another that also use information from many-to-one matching. The dataset of paiwise stable matches is simulated from a model of the form

$$\begin{aligned} v_i &= x_i \alpha + \varepsilon_i \\ u_j &= z_j \beta + \eta_j \end{aligned}$$

where $x_i,\ z_j,\ \varepsilon_i,\ \eta_j$ are distributed as standard normal random variables. We use $J = 500$ firms, each firm j has capacity q_j drawn uniformly at random from $\{1, \ldots, 10\}$. The number of workers in the simulation is $N = \sum c_j$. A pairwise stable match $\mu : \{1, \ldots, N\} \to \{1, \ldots, J\}$ is computed for $\alpha = 1$ and $\beta = 1$. Using the same draw of observables and firm capacities, the variables ε_i and η_j are simulated $S = 1000$ times, and a pairwise stable match μ_s^θ can be computed for each $s \in \{1, \ldots, S\}$ as a function of $\theta = (\alpha, \beta)$. We then compute two sets of moments

$$\hat{\psi}_{ov} = \frac{1}{N} \sum_i x_i z_{\mu(i)} \tag{3.4a}$$

$$\hat{\psi}_{ov}^S (\theta) = \frac{1}{B} \sum_b \frac{1}{N} \sum_i x_i z_{\mu_b^\theta (i)} \tag{3.4b}$$

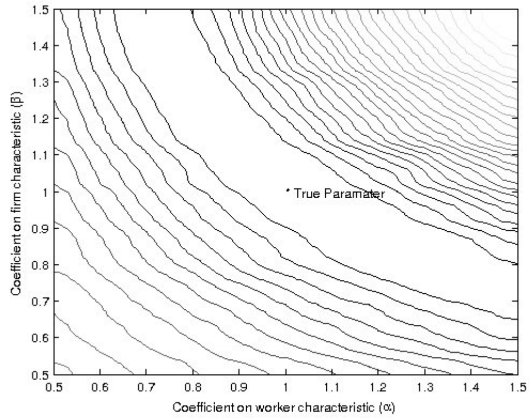
and

$$\hat{\psi}_w = \frac{1}{N} \sum_i \left(x_i - \frac{1}{\left| \mu^{-1} (\mu (i)) \right|} \sum_{i' \in \mu^{-1}(\mu(i))} x_{i'} \right)^2 \tag{3.5a}$$

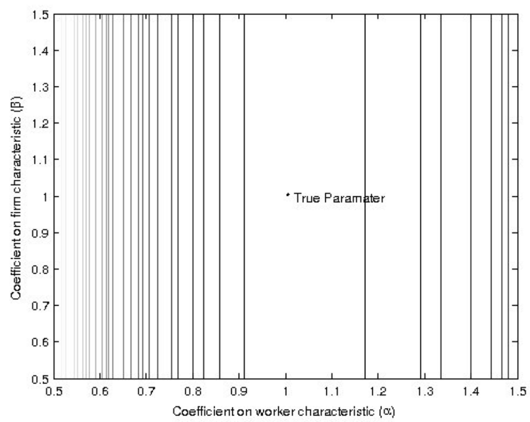
$$\hat{\psi}_w^S (\theta) = \frac{1}{B} \sum_b \frac{1}{N} \sum_i \left(x_i - \frac{1}{\left| \left(\mu_b^\theta\right)^{-1} \left(\mu_b^\theta (i)\right) \right|} \sum_{i' \in \left(\mu_b^\theta\right)^{-1} \left(\mu_b^\theta (i)\right)} x_{i'} \right)^2 . \tag{3.5b}$$

The first set, $\hat{\psi}_{ov}$ and $\hat{\psi}_{ov}^S (\theta)$, captures the degree of assortativity between the characteristics x and z in the pairwise stable matches in the generated data, and as a function of θ. For a given $\alpha > 0$ (likewise $\beta > 0$), this covariance should be increasing in β (likewise α). The second set, $\hat{\psi}_w$ and $\hat{\psi}_w^S (\theta)$ capture the within firm variation in the characteristic x. Our identification argument suggests that for larger values of α will result in lower values of $\hat{\psi}_{ov}^S (\theta)$. Using these moments, we construct an objective function $\hat{Q}(\theta) = \left\| \hat{\psi} - \hat{\psi}^S (\theta) \right\|_W$ where $\hat{\psi} = \left(\hat{\psi}_{ov}, \hat{\psi}_w \right)'$, $\hat{\psi}^S (\theta) = \left(\hat{\psi}_{ov}^S (\theta), \hat{\psi}_w^S (\theta) \right)'$ and W indexes the norm.

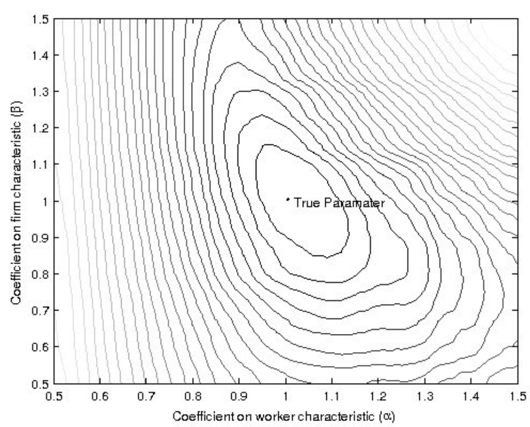
Figure 3.1(a) presents a contour plot of an objective function that only penalizes deviations of $\hat{\psi}_{ov}$ from $\hat{\psi}_{ov}^S (\theta)$. This objective function only using information in the sorting



(a) Sorting Moments Only



(b) Within-program Moments Only



(c) All Moments

Figure 3.1: Importance of Many-to-one Matches: Objective Function Contours

between x and z to differentiate values of θ. We see that pairs of parameters, α and β, with large values of α and small values of β yield identical values of the objective function. These contour sets result from identical values of $\hat{\psi}_{ov}^{S}(\theta)$, illustrating that this moment cannot distinguish between values along this set. In particular, the figure shows that the objective function has a trough containing the true parameter vector with many values of θ yielding similar values of the objective function.

In Figure 3.1(b), we consider an objective function that only penalizes deviations of $\hat{\psi}_w$ from $\hat{\psi}_w^S(\theta)$. The vertical contours indicate that the moment is able to clearly distinguish values of α because the moment $\hat{\psi}_w^S(\theta)$ is strictly decreasing in α. However, the shape of the objective function indicates that this moment cannot distinguish different values of β.

Finally, the plots of an objective function that penalizes deviations from both $\hat{m}_w$ and $\hat{m}_{ov}$ (Figure 3.1(c)) show that we can combine information from both sets of moments to identify the true parameter. This objective function has a unique minimum, close to the true parameter. Together, Figures 3.1(a)-(c) illustrate the importance of using both these types of moments in estimating our model.

3.4 Estimation

We consider estimation for a parametric class of models in which the functions determining the latent utilities of workers for firms and vice-versa are known up to a finite dimensional parameter $\theta \in \Theta \subseteq \mathbb{R}^{K_\theta}$. We assume utilities are generated by

$$\begin{aligned} u(z, \eta; \theta) &= g(z; \theta) + \eta \\ v(x, \varepsilon; \theta) &= h(x; \theta) + \varepsilon \end{aligned}$$

where $g : \zeta \times \Theta \to \mathbb{R}$ and $h : \chi \times \Theta \to \mathbb{R}$ are known-functions that are Lipschitz-continuous in each of their arguments. We assume that the marginal distributions m^{ε} and m^{η} admit known densities f_ε and f_η.

Our results are for a sample of J firms, each with $\bar{c} = 2$ slots each, and consider the properties of the estimator as $J \to \infty$. The number of workers is $N = \bar{c}J$. The characteristics of each worker are sampled from the measure $m^{x,\varepsilon}$ and the characteristics of the firm are sampled from $m^{z,\eta}$. Instead of considering a sampling process in which pairs

(x_i, ε_i) and (z_j, η_j) are drawn, it will be convenient to first sample N and J draws from m^v and m^u respectively, and then sample $x_i|v_i$ and $z_j|u_j$ from their respective conditional distributions. This sampling process has an identical distribution for (x_i, ε_i) and (z_j, η_j) as sampling directly from $m^{x,\varepsilon}$ and $m^{z,\eta}$ directly.

We will study the estimator defined by

$$\theta_N = \arg\min_{\theta \in \Theta} \|\psi_N - \psi_N(\theta)\|_W, \tag{3.6}$$

where ψ_N are finite-dimensional moments computed from the sample, $\psi_N(\theta)$ are computed from the observed sample of firms and workers as a function of θ, and W defines a norm. For instance, we may use a quadratic form in the difference $(\psi_N - \psi_N(\theta))$.

Let m^u, m^v be the image measures of $m^{z,\eta}$ and $m^{x,\varepsilon}$ under $u(z, \eta; \theta_0)$ and $v(x, \varepsilon; \theta_0)$ respectively, and m_N^u, m_J^v be their empirical analogues. We will treat these densities as known functions. Let $\Psi : \chi \times \chi \times \zeta \to \mathbb{R}^{K_\Psi}$ be a moment function. We make the following regularity assumptions on these primitives.

Assumption 3.5. *(i) $\Psi(x_1, x_2, z)$ is bounded with bounded partial derivatives and symmetric in x_1 and x_2*

(ii) The densities f_ε and f_η are bounded and have bounded derivatives

(iii) The densities f_ε and f_η have full support on the real line

(iv) m^x and m^z admit densities f_X and f_Z

(v) The conditional densities $f_{X|v}(x)$ and $f_{Z|v}(z)$ have uniformly bounded derivatives.

The pairwise stable match can then be computed for such a sample by matching workers and firms on quantiles. The data thus consist of $N = 2J$ matches, $\{(x_{2j-1}, x_{2j}, z_j)\}_{j=1}^J$. We can construct empirical moments of the form

$$\psi_N = \frac{1}{N} \sum_j \Psi(x_{2j-1}, x_{2j}, z_j). \tag{3.7}$$

where Ψ is symmetric in its first two arguments. The moments in equations (3.4a) and (3.5a) can be written in this form when we have a dataset with two-to-one matching.

In the population dataset, firms with the q-th quantile of m^v are matched with workers on the q-th quantile of m^u. The moment can be written as

$$\psi = \int_0^1 \tilde{\psi}\left(F_v^{-1}(q), F_v^{-1}(q), F_u^{-1}(q); \theta_0\right) dq$$

where F_v and F_u are the cdf corresponding to m^v and m^u respectively and $\tilde{\psi}(v_1, v_2, u)$ is the expectation of $\Psi(X_1, X_2, Z)$ given that X_1 and X_2 are drawn from $m^{x|v}$ and Z is drawn from $m^{z|u}$. The term $\psi(v_1, v_2, u)$ is given by

$$\tilde{\psi}(v_1, v_2, u; \theta) = \frac{\int \Psi(X_1, X_2, Z) f_\varepsilon(v - h(X_1; \theta)) f_\varepsilon(v - h(X_2; \theta)) f_\eta(u - g(Z; \theta)) dm^x dm^x dm^z}{\int f_\varepsilon(v - h(X_1; \theta)) f_\varepsilon(v - h(X_2; \theta)) f_\eta(u - g(Z; \theta)) dm^x dm^x dm^z}. \tag{3.8}$$

Our first result shows that the empirical moments converge at the true parameter θ_0.

Proposition 3.7. *Let ψ^k and ψ_N^k denote the k-th dimensions of ψ and ψ_N respectively. If Assumption 3.5 is satisfied, then for each $k \in \{1, \ldots, K_\Psi\}$, $\psi_N^k - \psi^k$ converges in probability to* 0.

Proof. See Appendix C.2.2. □

The primary technical difficulty arises from the dependent data nature of the observed matches. By using the sampling fiction in which the utilities u and v are drawn first, we can condition on the utility-matches in the data. The observed characteristics of the matched agents are then sampled conditional on this utility draw. Since the distribution of utilities converge to the population distribution, this sampling process, although identical to drawing the characteristics directly from $m^{x,\varepsilon}$ and $m^{z,\eta}$, allows a more tractable approach to proving consistency of the moments. The proof technique is based on leveraging the triangular array structure implied by this process.

For estimation, we also need to consider the population and empirical analogs of ψ evaluated at values of θ other than θ_0. The population analog is given by

$$\psi(\theta) = \int_0^1 \tilde{\psi}\left(F_{v;\theta}^{-1}(q), F_{v;\theta}^{-1}(q), F_{u;\theta}^{-1}(q); \theta\right) dq$$

where $F_{v;\theta}(v) = \int_{-\infty}^{v} F_\varepsilon(v - h(X; \theta)) dm^x$, $F_{u;\theta}(u) = \int_{-\infty}^{u} F_\eta(u - g(Z; \theta)) dm^z$ and $\tilde{\psi}$ is defined in equation (3.8) above. We study an estimator that uses the following sample analog of $\psi(\theta)$ as a function of θ,

$$\psi_N(\theta) = \int_0^1 \tilde{\psi}\left(F_{N,v;\theta}^{-1}(q), F_{N,v;\theta}^{-1}(q), F_{J,u;\theta}^{-1}(q); \theta_0\right) dq \tag{3.9}$$

$$\tilde{\psi}_N(v_1, v_2, u; \theta) = \frac{\int \Psi(X_1, X_2, Z) f_\varepsilon(v_1 - h(X;\theta)) f_\varepsilon(v_2 - h(X;\theta)) f_\eta(u - g(Z;\theta)) dm_N^x dm_N^x dm_J^z}{\int f_\varepsilon(v_1 - h(X;\theta)) f_\varepsilon(v_2 - h(X;\theta)) f_\eta(u - g(Z;\theta)) dm_N^x dm_N^x dm_J^z},$$

where $F_{N,v;\theta}$ and $F_{J,u;\theta}$ are empirical cdf functions from a random sample from $F_{v;\theta,m_N^x}(v) = \int_{-\infty}^{v} F_\varepsilon(v - h(X;\theta)) dm_N^x$ and $F_{v;\theta,m_J^z}(v) = \int_{-\infty}^{u} F_\eta(u - g(Z;\theta)) dm_J^z$ respectively. $\psi_N(\theta)$ can be computed by first generating a simulated sample of ε and η to simulate $F_{N,v;\theta}$ and $F_{J,u;\theta}$, and then using the expression in equation (3.9) to compute $\psi_N(\theta)$. It may also be possible to create a simulation analog of $\psi_N(\theta)$, that uses a second simulation step to approximate the integral. More specifically, we may independently sample from the conditional distributions of X and Z given the measures m_N^x and m_J^z and simulated values of v_i and u_j.

The next result proves uniform convergence of the difference $\psi(\theta) - \psi_N(\theta; m_N^x, m_J^z)$, a result required for consistency of the estimator.

Proposition 3.8. *Let $\psi^k(\theta)$ and $\psi_N^k(\theta; m_N^x, m_J^z)$ denote the k-th dimensions of $\psi(\theta)$ and $\psi_N(\theta; m_N^x, m_J^z)$ respectively. If Assumptions 3.5(i) - (iii) are satisfied, then for each $k \in \{1, \ldots, K_\Psi\}$, $\left|\psi^k(\theta) - \psi_N^k(\theta; m_N^x, m_J^z)\right|$ converges in outer probability to 0 uniformly in θ.*

Proof. See Appendix C.2.3. □

Again, the proof leverages the triangular sampling structure, conditioning on the drawn utilities. The first step is to prove that the cumulative distribution of utilities converge uniformly in θ as the sampled observed characteristics, m_N^x and m_J^z converge to their population analogs. Given this, we can take advantage of the triangular structure to construct the expected value of Ψ given an empirical distribution of sampled utilities by computing $\tilde{\psi}_N(v_1, v_2, u; \theta)$ along the quantiles of $F_{N,v;\theta}$ and $F_{J,u;\theta}$.

The proof is not a direct extension of techniques in Proposition 3.7. Intuitively, while the expectation Ψ given the empirical observable measures m_N^x and m_J^z converge uniformly in θ, the particular triples (X_1, X_2, Z) that are matched are less tractable across values of θ. The second stage of sampling, conditional on the sampled utilities implicit in equation (3.9) lends tractability to $\tilde{\psi}_N$, hence $\psi_N(\theta)$.

Finally, we use the following standard assumptions to prove consistency of the estimator defined in equation (3.6).

Assumption 3.6. *(i) The parameter space Θ is compact*

(ii) There is a unique $\theta_0 \in \Theta$ for which

$$\psi(\theta) = \psi(\theta_0)$$

(iii) For the $\|\cdot\|_W$ and for any $\varepsilon > 0$, there is $\delta > 0$ such that $\|\psi(\theta) - \psi(\theta_0)\|_W < \delta \Rightarrow \|\theta - \theta_0\| < \varepsilon$.

(iii) The norm $\|\cdot\|_W$ is continuous in its argument

Theorem 3.2. *Let $\widehat{\theta}_N = \arg\min_{\theta\in\Theta} \|\psi_N - \psi_N(\theta)\|_W$. If Assumptions 3.5 and 3.6 are satisfied, then $\widehat{\theta}_N$ converges in probability to θ_0.*

Proof. See Appendix C.2.4. □

3.5 Monte Carlo Evidence

This section presents Monte Carlo evidence of a simulation based estimator from synthetic datasets of varying size and models of varying complexity to assess the properties of a method of simulated moments estimator. The results are presented for a simulation based estimator of the form

$$\begin{aligned} \hat{\theta}_N &= \arg\min_{\theta\in\Theta} \|\psi_N - \psi_{N,S}(\theta)\|_W & (3.10)\\ &= \arg\min_{\theta\in\Theta} \left[(\psi_N - \psi_{N,S}(\theta))' W (\psi_N - \psi_{N,S}(\theta))\right]^{1/2} & (3.11) \end{aligned}$$

where ψ_N is as defined in equation (3.7) and $\psi_{NS}(\theta)$ is computed identically to $\psi_{NS}(\theta)$ using $S = 100$ simulations. For each simulation s, we sample the unobservables ε_i and η_j, compute the unique pairwise stable match and compute $\psi_{N,s}(\theta)$ for the simulated matches. The quantity $\psi_{N,S}(\theta) = \frac{1}{S}\sum_s \psi_{N,s}(\theta)$. The moments used are as defined in equations (3.4a) and (3.5a). One within moment is included for each observed component of x. The overall moments include each component of x interacted with each component of z.

We begin by assessing the performance of the estimator for the double-vertical model for which the previous sections of the paper presents non-parametric identification results and limit theorems. We also present Monte Carlo evidence on models with workers having heterogeneous preferences for firms although we do not have formal theory on those models.

3.5.1 Design of Monte Carlo Experiments

Our Monte Carlo experiments vary the number of programs, $J \in \{100, 500\}$, and the maximum number of residents matched with each program $\bar{c} \in \{5, 10\}$. For each program j, the capacity c_j is chosen uniformly at random from $\{1, \ldots, \bar{c}\}$. The number of residents is a random variable set at $N = \sum c_j$. We will use up to three characteristics for residents and up to four characteristics for programs. The characteristics z_j of program j are distributed as

$$z_j = (z_{j1}, z_{j2}, z_{j3}) \sim N(a, I_3)$$

where $a = (1, 2, 3, 4)$ and I_4 is a 4×4 identity matrix. The characteristics of the residents, x_i are distributed as

$$x_i = (x_{i1}, x_{i2}, x_{i3}, x_{i4}) \sim N(b, I_4)$$

where $b = (1, 2, 3)$ and I_3 is a 3×3 identity matrix.

For each model specification, we generate 500 samples indexed by b and parameter estimates $\hat{\theta}_b$.[4] The confidence intervals are generated by using a parametric bootstrap described in Appendix C.3.

[4]The $b-$th (pseudo-random) sample is generated from a Mersenne Twister algorithm with the seed b.

3.5.2 Results

The Double Vertical Model

We present Monte Carlo evidence from a model with no preference heterogeneity. The preferences are of the form,

$$v_i = x_i\alpha + \varepsilon_i \tag{3.12a}$$

$$u_j = z_j\beta + \eta_j \tag{3.12b}$$

where $\varepsilon_i \sim N(0,1)$ and $\eta_j \sim N(0,1)$. Table 3.1 presents results from two specifications. The specification in Column (1) has a single observable characteristic on each side of the market and column (2) has two observable characteristics. With few exceptions, the bias, the root mean squared error (RMSE) and the standard error fall with J and $\bar{q}$ for both specifications. The coverage ratios of 95% confidence intervals constructed from the proposed bootstrap approximation are mostly between 90% and 98%, particularly for simulations with a larger sample sizes, particularly for estimates of α. Also notice that estimates for α are more precise than estimates of β in both specfciations and all sample sizes.

Heterogeneous Preferences

Preference models without heterogeneity may be quite restrictive for some empirical applications. While we do not have formal results on identification or estimation of models with preference heterogeneity, we present Monte Carlo evidence from a model in which workers have heterogeneous preferences for firms. We conduct Monte Carlo simulations for pairwise stable matches using preference models of the form

$$v_i = x_i\alpha + \varepsilon_i \tag{3.13}$$

$$u_{ij} = z_j\beta + \sum_{k,l}\gamma_{kl} \times x_{i,k} \times z_{j,l} + \eta_j \tag{3.14}$$

where $\varepsilon_i \sim N(0,1)$ and $\eta_j \sim N(0,1)$. In this model, workers have varying preferences for the firm characteristic z based on their characteristic x. Table 3.2 presents results from two specifications, one with one interaction term and another with two interactions. As in the

Table 3.1: Monte Carlo Evidence: Double-Vertical Model

	One Characteristic (1)		Two Characteristics (2)			
	$\alpha_1\,(x_1)$	$\beta_1\,(z_1)$	$\alpha_1\,(x_1)$	$\alpha_2\,(x_2)$	$\beta_1\,(z_1)$	$\beta_2\,(z_2)$
$J = 100,\ \bar{q} = 5$						
True Par.	1	1	1	2	-1	2
Bias	0.005	0.053	0.033	0.042	-0.593	1.134
RMSE	0.093	0.239	0.270	0.412	1.063	1.938
SE	0.131	0.403	0.151	0.225	0.419	0.803
Coverage	0.954	0.970	0.868	0.850	0.760	0.750
$J = 100,\ \bar{q} = 10$						
True Par.	1	1	1	2	-1	2
Bias	0.002	0.046	-0.028	-0.036	-0.478	0.970
RMSE	0.063	0.196	0.143	0.211	0.932	1.703
SE	0.073	0.341	0.119	0.166	0.392	0.768
Coverage	0.972	0.978	0.894	0.914	0.800	0.820
$J = 500,\ \bar{q} = 5$						
True Par.	1	1	1	2	-1	2
Bias	0.000	0.002	-0.018	-0.022	-0.066	0.134
RMSE	0.042	0.086	0.052	0.069	0.207	0.383
SE	0.057	0.153	0.062	0.093	0.093	0.172
Coverage	0.934	0.978	0.950	0.954	0.808	0.848
$J = 500,\ \bar{q} = 10$						
True Par.	1	1	1	2	-1	2
Bias	0.000	0.004	-0.012	-0.017	-0.043	0.088
RMSE	0.027	0.080	0.037	0.049	0.159	0.296
SE	0.033	0.141	0.055	0.076	0.094	0.177
Coverage	0.968	0.992	0.966	0.976	0.888	0.894

model with no preference heterogeneity, the bias, root mean square error falls with J and $\bar{c}$, and the coverage ratios are close to correct. The result suggests that a simulation based estimator for the model with preference heterogeneity may have desirable large sample properties as well.

3.6 Conclusion

This paper provides the first results on the indentification and estimation of preferences from data from a matching market described by pariwise stability and non-transferable utility, when data only on final matches are observed. Our results are restricted to the case when preferences on both sides are homogeneous. We show that using information available in many-to-one matching is necessary and sufficient for non-parametric identification if data on a single large market is observed. We also prove consistency of an estimator for a parametric class of models. Finally, we present Monte Carlo evidence on a simulation based estimator.

There are several avenues for future research on both identification and estimation for similar data environments. While we show that it is necessary to use information from many-to-one matching for identification with data on a single large market, it may be possible to prove identification results using variation in the characteristics of market participants. This can be particularly important for the emprical study of marriage markets in the non-transferable utility framework. Our formal results are also restricted to the case with homogeneous preferences on both sides of the market. Extending this domain of preferences is particularly important. A treatment of heterogeneous preferences on both sides of the market may be of particular interest, but may need to confront difficulties arising from the multiplicity of equilibria. Finally, we have also left asymptotic theory for the estimator proposed in this paper and the exploration of computationally more tractable estimators for future research.

Table 3.2: Monte Carlo Evidence: Observable Heterogeneity

	One Interaction (1)			Two Interactions (2)					
	$\alpha_1\,(x_1)$	$\beta_1\,(z_1)$	$\gamma_1\,(z_1 \times x_3)$	$\alpha_1\,(x_1)$	$\alpha_2\,(x_2)$	$\beta_1\,(z_1)$	$\beta_2\,(z_2)$	$\gamma_{1,3}\,(z_1 \times x_3)$	$\gamma_{3,3}\,(z_3 \times x_3)$
$J = 100,\ \bar{q} = 5$									
True Par.	1	1	1	1	2	-1	2	-1	2
Bias	0.013	-0.196	0.195	0.027	0.030	-0.090	0.184	-0.112	0.175
RMSE	0.086	1.335	1.066	0.201	0.332	0.396	0.462	0.292	0.405
SE	0.136	1.861	1.389	0.114	0.175	0.219	0.293	0.179	0.284
Coverage	0.988	0.950	0.944	0.934	0.908	0.802	0.854	0.862	0.906
$J = 100,\ \bar{q} = 10$									
True Par.	1	1	1	1	2	-1	2	-1	2
Bias	0.005	-0.204	0.158	-0.005	-0.003	-0.067	0.110	-0.067	0.104
RMSE	0.061	1.002	0.765	0.114	0.187	0.317	0.378	0.235	0.290
SE	0.086	1.539	1.129	0.103	0.146	0.218	0.290	0.175	0.278
Coverage	0.984	0.940	0.964	0.960	0.966	0.884	0.912	0.906	0.964
$J = 500,\ \bar{q} = 5$									
True Par.	1	1	1	1	2	-1	2	-1	2
Bias	-0.001	0.025	0.007	-0.015	-0.018	-0.012	0.028	-0.018	0.026
RMSE	0.039	0.666	0.371	0.040	0.059	0.107	0.131	0.074	0.085
SE	0.056	0.885	0.498	0.056	0.082	0.102	0.150	0.089	0.145
Coverage	0.978	0.926	0.960	0.962	0.984	0.956	0.978	0.978	0.994
$J = 500,\ \bar{q} = 10$									
True Par.	1	1	1	1	2	-1	2	-1	2
Bias	0.000	-0.043	0.032	-0.008	-0.011	-0.010	0.021	-0.014	0.020
RMSE	0.027	0.559	0.302	0.026	0.038	0.099	0.120	0.074	0.081
SE	0.035	0.750	0.408	0.052	0.072	0.103	0.154	0.089	0.139
Coverage	0.976	0.952	0.988	0.986	0.988	0.962	0.984	0.970	0.990

Appendix A

Appendices to Chapter 1

A.1 Data Appendix

This appendix provides details about our data sources and the steps behind our sample construction.

A.1.1 Data Sources

Office of Student Enrollment and Planning Operations (OSEPO), which runs high school admissions maintains the High School Application Processing System (HSAPS) file. The HSAPS file contains the New York City Public High School Admissions Application of all 8th graders and a smaller number of 9th graders. The file contains students' choice schools in preference order, priority information for each school, and the high school assignments at the end of the first round of the match and at the end of the main round. The OSEPO student file provides school attended, home street address, grade, gender, ethnicity. In separate files, the department also provided us with additional student characteristics such as scores on middle school standardized tests, limited English proficiency status, special ed status.

Students are indexed by their NYCDOE ID number in each file, which we used to merge the files together. Each eight-grade non-private middle school student in the OSEPO student file could be merged uniquely with a student in the HSAPS file. Less than 0.45%

of students that appeared with known assignments in the HSAPS file could not be merged with a student in the OSEPO student file. These students were not included in the analysis.

We append the data with distance between each student's home address in the application file and the address of each school calculated by Microsoft Mappoint. Corrections to the addresses, when necessary, were made using Google Geocoder followed by manual checks and corrections to ensure data quality. We have also used the 2000 US Census to obtain block group level income demographics. Some fields are not available at the block group level in the Census, so in these cases, we use the tract level data.

Data on schools were taken from the 2003-04 report card files provided by NYCDOE, which contain information on school enrollment statistics, racial composition of student body, attendance rates, suspensions, teacher numbers and experience, and student achievement of the graduating class. A unique identifier for each school is used to merge the school data with other files. We collected the data on programs characteristics from High School Directories published by NYCDOE and made available to students before the application process.

A.1.2 Student Samples

We work with two *welfare* samples from which we measure final assignments and a *demand* sample from which we estimate preferences. Our aim is to work with the largest set of first-time applicants to non-Specialized public high schools that live in New York City and matriculate into non-Specialized public high schools in 2002-03 and in 2003-04. To this end, the welfare samples for 2002-03 and 2003-04 are obtained from the entire student universe in the OSEPO student files. Table A.1 summarizes the steps involved in selecting the sample.

Welfare Sample

Columns (1) and (2) in Table A.1 summarize the selection of welfare samples for the old mechanism in 2002-03 and new mechanism in 2003-04, respectively. We start with the NYC Department of Education's records in the OSEPO files, which contain 100,669 records for 2002-03. Excluding ninth graders applying to 10th grade programs, we obtain

Table A.1: Student Sample Selection

Student Samples:	Mechanism Comparison		Demand Sample
	Old Mechanism	New Mechanism	New Mechanism
	(1)	(2)	(3)
Total Number of Students in the NYC DOE file	100669	97569	
Total Number of Students in the rank data			87355
Excluding students in both 0203 and 0304 files from 0304		96275	86608
Excluding ninth graders	92623	89062	81297
Excluding private middle school students	80833	78183	71996
Excluding with addresses outside the five boroughs	80725	78089	71916
Total number of students with known assignments to inside options	75515	73989	
Excluding students attending specialized high schools	72725	70992	
Excluding students attending charter schools	72681	70886	
Excluding students in closed schools, and new schools	70358	66921	
Excluding top 2% students			70123
Excluding students that did not rank any inside options			69907

Notes: The top 2% indicator appears to be different across years. A student has invalid census information if address is missing, cannot be geocoded or places the student outside of New York City. A distance observation is invalid if it is missing or is greater than 65 miles.

92,623 records for all students required to enroll in ninth grade at a high school in the academic years 2002-03.

As our choice set in the demand analysis will be restricted to unspecialized, non-charter high schools in the public school system, we do not include students that matriculated to such schools in the welfare sample. Column (1) reports the number of remaining students after excluding students applying from private middle schools, students applying from Westchester and Long Island, students without known assignments to NYC public high schools, students attending specialized high schools, students attending charter schools, students with invalid census data and students with invalid distance observations. A total of 827 students or 1.16% of the 2003-2004 sample were also present in 2002-03 presumably because these students repeated eighth grade. These students were considered a part of the 2002-03 sample and only their 2002-03 assignment into high school is considered in our analysis. We also eliminate students who are assigned to schools that were not in the initial high school directory, and either closed before the start of admissions cycle or opened after it started. The directory of New Public High Schools for the 2003-04 school year was released online on March 10, 2004, more than 5 months after admissions cycle.

These sample selection criteria leave us with 69,100 students in 2002-03 and 66,466 students in 2003-04. Students that may have been assigned to a high school program through a process other than the main round are included in these samples.

Demand Sample

Since we do not have access to choice data in 2002-03, there is no demand sample for that year. Column (3) in Table A.1 summarize the selection of the demand samples for the new mechanism in 2003-04. The demand sample is obtained from the HSAPS file in 2003-04. Due to the incentive issues discussed in the text, we use data only from the main round of the mechanism as this round has the most desirable incentive properties.

In order to most closely match the construction of the welfare sample without causing selection problems, we select the demand sample only on characteristics that can be considered as exogenous at the time of participation. Column (3) reports the number of remaining students after excluding ninth graders, students applying from private middle schools, top-2-percent students, students applying from Westchester and Long Island, students with invalid census data and students with invalid distance observations, students that appear in the 2003-04 HSAPS file and students who did not rank any inside options. We do not exclude students without known assignments to NYC public high schools, students attending specialized high schools, students attending charter schools in order to avoid selecting on choice to leave the public school system. Due to an apparent discrepancy between the top-2-percent indicators across years, top-2-percent students are not excluded in the welfare samples. We also include students who may have ranked a closed or new school to measure preferences, even though students assigned to these programs do not contribute to the welfare sample.

These selections into the sample leave us with 69,582 students that we use for the demand analysis.

Test Scores

There are several standardized tests taken by middle school students in NYC. To avoid the concern that two different tests may not be comparable indicators of student achievement, we identify the modal standardized tests in Math and reading taken by students in our sample. These are the May tests with codes CTB and TEM respectively. Of the students that did not take either of these tests in May, at most 10% (¡2% of the full sample) took a different standardized test in the same subject while in middle school. The distribution and support of the test scores were verified to be similar across the two years in our samples. Some students took the test multiple times. The highest score obtained by a student was used in these instances.

In 2002-03, the math and reading scores were missing for 13.56% and 17.55% students from our final sample respectively. For the 2003-04 welfare sample, records were missing for 8.29% and 13.57% students respectively for math and reading. In the demand sample the corresponding fractions were 7.13% and 12.56%.

A.1.3 Programs/Schools

We consider assignment into all public high schools in New York City that are not chartered, specialized or parochial. Our analysis uses two samples of schools, one for each year in our analysis. Table A.2 summarizes our selection rules for the samples.

To construct these samples, we started with the set of schools and programs in the HSAPS file. We added the set of school programs that were ranked by any student in our demand sample. This initial set consists of 743 programs and 301 schools in 2002-03 and 677 programs and 293 schools in 2003-04.

In 2003-04, this list contained 62 small, parochial school programs. We verified that each of the 130 students matriculating to these school programs were private middle-schoolers. We dropped these schools since we do not study private middle-schoolers. Subsequently, we dropped all charter and specialized high school programs. We also dropped other school programs which do not have assignments and were not ranked by any student in our demand sample.

Finally, 14 continuing student programs accepted students only from their associated middle school. As these programs cannot be chosen by students that were not in that school in eight grade, the analysis of the choice data requires that a rank order list submitted by a student that does not contain one of these programs is not interpreted as stating a dislike for the program. Hence, these programs were combined with a generic program (unscreened, English, general/humanities/math; see below). Rank order lists for students that ranked both the continuing students only program and the associated program were modified as described below.

School Characteristics

The school characteristics were taken from NYCDOE report card files. The files provide information on a school's enrollment statistics, racial composition of student body, attendance rates, suspensions, teacher numbers and experience, and student achievement of the graduating class. A separate NYCDOE file provide data on the school addresses. A unique identifier for each school allows the data to be merged across files. Across the two years, the school identifiers in the files were inconsistent for a small number of schools in our sample. These were matched by name and address of the school. One school moved from Brooklyn to New York and was investigated to ensure that the records were appropriately matched.

There were significant differences in the file formats and field names across the years

Table A.2: Selection of the School Sample

	Old Mechanism		New Mechanism	
	Programs	Schools	Programs	Schools
	(1)	(2)	(3)	(4)
A. Sample Selection Process				
Programs with assignments recorded by the NYC public schools	743	301	669	293
Adding programs that were ranked but had no assignments			677	294
Excluding parochial schools	681	239	677	294
Excluding specialized schools	669	232	665	287
Excluding charter schools	667	230	663	285
Excluding programs with no assignments or rankings	637	225	648	284
B. Description of Final Sample				
Programs/schools with assignments from our student sample	637	225	639	284
Numbers excluding new and closed schools	612	215	558	235
Programs/schools ranked by students in our sample			497	234

Notes. 13 continuining student programs were merged with a generic program in the host school. The parochial schools in 2002-03 only in the report card files. In order to keep the school characteristics constant across years, we use the data from the 2003-04 report cards as the primary source. Except for data on the Math and English achievements, variable descriptions were comparable across years. For these comparable variables, we used the 2002-03 data only when the 2003-04 data were not available. Table A.3 reports the coverage of the characteristics for the school samples.

Table A.3: Coverage of School Characteristcs

	2002-2003 (1)	2003-2004 (2)	Schools in Both Years (3)
Total number of schools in the sample	215	234	215
9th grade enrollment	196	199	189
10th grade enrollment	179	175	172
11th grade enrollment	155	151	148
12th grade enrollment	154	150	147
Race	196	199	189
Attendance Rate	196	199	189
Percent Suspensions	187	186	180
Percent Free Lunch	196	198	189
Percent of teachers less than 2 years	219	223	212
Students with math in the 85-100% range	198	200	191
Students with english in the 85-100% range	180	177	173
Total graduates	172	169	166
Regents diplomas	171	168	165
Percent attending college	171	167	165
Student teacher ratio	196	199	189
Expenditure	172	168	165
Teachers	196	199	189
Capacity	194	196	187

Notes: Table reports the fraction of schools with the characteristic in the old and new mechanism.

Program Classification

The program characteristics are taken from the official NYC High School Directory that is made available to students before the application process. Reliable data on program types was not available in 2002-03. For that year, the program types were imputed from the 2003-04 program types if the program was present in both years. Otherwise, the program was categorized as a general program.

There are a very large number of program types. We aggregated them into fewer broad categories. The following list explains our aggregation rules. For example, a Dance program and a Vocal Performance program are both classified as a Arts program.

1. *Arts:* Dance, Instrument Performance, Musical Theater, Performing and Visual, Performing Arts, Theater, Theater Tech, Visual Arts, Vocal Performance.

2. *Humanities/Interdisciplinary*: Education, Humanities/Interdisciplinary.

3. *Business/Accounting*: Accounting, Business, Business Law, Computer Business, Finance, International Business, Marketing, Travel Business.

4. *Math/Science*: Math and Science, Science and Math, Engineering, Engineering – Aerospace, Engineering – Electrical, Environmental.

5. *Career*: Architecture, Computer Tech, Computerized Mech, Cosmetology, Journalism, Veterinary, Vision Care Technology.

6. *Vocational*: Auto, Aviation, Clerical, Construction, Electrical Construction, Health, Heating, Hospitality, Plumbing, Transportation.

7. *Government/law*: Law, Law Enforcement, Law and Social Justice, Public Service.

8. *Other:* Communication, Expeditionary, Preservation, Sports.

9. *Zoned*

10. *General*: General, Unknown.

Finally, some programs adopt a language of instruction other than English. We categorized the languages into Spanish, English, Asian Languages and Other.

Program Capacities

Program capacities are not provided separately in the data files. We have estimated program capacities from the actual match files and students' September 2004 school assignments. The capacity of each program is initially set to zero. If a student in our demand sample is assigned a program in September, the capacity of the program is increased by one. Otherwise, if the student is assigned a program in the main round the capacity of the program is increased by one. Finally, if a student is not assigned in the main round and assigned a program in the supplementary round, the capacity of the program is increased by one.

EdOpt programs are divided into six buckets, High Select, High Random, Middle Select, Middle Random, Low Select and Low Random. The bucket capacities are calculated as above by taking into account the category of the assigned student. For example, if a student of High category is assigned an EdOpt program, then the capacity of a High bucket is increased by one. If the current capacity of the High Select bucket is less than or equal to that of High Random, then the capacity of the High Select bucket is increased, otherwise the capacity of the High Random bucket is increased.

Program Priorities

The data contains admissions criteria for each program in the samples. The assignment following fields determine a student's priority order at programs. Priority group is a number assigned by NYCDOE depending on students' home addresses and location of programs etc. High school rank is a number assigned by each program. This may reflect a student's ranking among all applicants to an EdOpt program, or whether a student attended the information session of a limited unscreened program, etc. These fields are provided for every student at every program that the student ranked. Students applying to an educational option program are categorized into one of three categories based on their score on the seventh grade standardized reading test: top 16 percent (High), middle 68 percent (Middle), and bottom 16 percent (Low). Student categories are provided in the HSAPS file.

Admissions criteria are explained in the text. An EdOpt program applies its admission criteria for each of its six buckets, High Select, High Random, Middle Select, Middle Random, Low Select and Low Random, as follows: A high bucket orders high category students first, then middle category students, then low category students. A middle bucket orders middle category students first, then high category students, then low category students. A low bucket orders low category students first, then high category students, then middle category students. A select bucket orders students within each category by priority order, then by high school rank. A random bucket orders students within each category by priority order.

A.1.4 Miscellaneous Issues

Modifications to the rank order list

1. Some students ranked a program that were either charter schools or specialized high schools in the main round. These programs are not in the sample of schools we consider and were likely ranked by the students in error. In such cases, the rank order lists were made contiguous where all programs ranked below a program not in the sample were moved up in the rank order lists. These programs were observed a total of 795 times in the data. Thirty students ranked only charter or specialized programs.

2. The rank order lists of students that ranked continuing student program needed modification. First, the lists of all students that ranked only the continuing student program were modified so that the student ranked the associated generic program instead. When students ranked both the generic program and the associated continuing student program, the list was modified so that only the associated program was ranked, and at the highest of the two ranked positions. All programs ranked at positions below the lower ranked of the two programs were moved up by one. A total of 46 students ranked both the continuing program and the generic program we mapped the continuing program to. In 17 cases, these ranks were not consecutive.

Table A.4: Choice Assigned in New Mechanism, by Length of Preferences

Choice	All	by Length of Preferences											
Assigned		1	2	3	4	5	6	7	8	9	10	11	12
Total	69907	4597	3282	4128	4622	4952	4776	4406	4390	4558	6135	9849	14212
1	31.9%	88.6%	40.7%	35.2%	31.9%	27.9%	28.6%	27.1%	25.7%	25.6%	25.4%	26.2%	25.2%
2	15.0%		39.8%	17.7%	15.1%	14.8%	14.6%	13.7%	13.9%	13.9%	15.2%	14.7%	14.6%
3	10.2%			24.3%	11.6%	11.6%	10.6%	10.0%	10.8%	9.9%	10.4%	10.4%	10.5%
4	7.3%				18.0%	9.3%	8.1%	7.9%	8.0%	7.6%	7.6%	7.8%	8.2%
5	5.4%					12.8%	7.0%	7.0%	6.3%	6.1%	6.6%	6.2%	6.7%
6	3.9%						10.2%	5.7%	4.9%	5.0%	4.9%	4.8%	5.3%
7	2.9%							8.1%	4.3%	4.4%	4.0%	4.1%	4.3%
8	2.0%								5.8%	3.4%	3.3%	2.9%	3.5%
9	1.5%									4.0%	2.8%	2.7%	2.8%
10	1.1%										3.2%	2.3%	2.6%
11	0.8%											2.6%	2.2%
12	0.5%												2.5%
Not Assigned	17.5%	11.4%	19.5%	22.8%	23.3%	23.6%	20.9%	20.6%	20.3%	20.1%	16.7%	15.3%	11.6%

Notes: This table reports choices assigned after the main round of 2003-04.

Table A.5: Assignment and Enrollment of Students in Centralized Mechanism

	Total	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th	12th
A. Students Offered Assignment in Main Round													
N	57658	4072	2641	3187	3545	3782	3776	3497	3499	3642	5113	8340	12564
Avg. rank of assign.	3.00	1.00	1.49	1.86	2.21	2.53	2.76	3.04	3.20	3.35	3.49	3.60	3.93
Std. rank of assign.	2.47	—	0.50	0.87	1.21	1.49	1.79	2.07	2.25	2.46	2.63	2.83	3.10
Accept assignment	92.7%	91.2%	88.5%	88.4%	90.2%	91.2%	92.3%	91.9%	93.0%	93.6%	94.5%	94.6%	94.3%
Enroll in pvt. school	2.5%	6.9%	7.4%	6.1%	4.5%	2.9%	2.4%	2.1%	1.9%	1.2%	1.0%	0.7%	1.0%
Remain in current sch.	1.2%	1.2%	2.0%	2.3%	1.9%	2.1%	1.4%	1.8%	1.4%	1.2%	0.9%	0.7%	0.6%
Attend spec. schools	0.1%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	0.2%	0.1%	0.1%	0.2%	0.1%	0.0%
Participate in round 3	0.3%	0.1%	0.2%	0.2%	0.4%	0.5%	0.6%	0.3%	0.6%	0.3%	0.3%	0.2%	0.3%
B. Students Unassigned After Main Round													
N	12249	525	641	941	1077	1170	1000	909	891	916	1022	1509	1648
Participate in round 3	52.6%	26.1%	44.8%	54.0%	54.1%	56.2%	55.6%	55.7%	52.7%	46.5%	43.5%	49.6%	68.2%
Enroll in Rd. 3 assign.	72.9%	73.0%	85.0%	76.0%	75.5%	77.8%	73.0%	75.9%	74.5%	68.8%	71.7%	69.5%	66.3%
Enroll in pvt. school	2.8%	6.7%	6.1%	4.7%	3.5%	3.8%	2.2%	1.7%	1.6%	1.9%	2.2%	1.4%	1.9%
Remain in current sch.	3.2%	6.7%	6.2%	5.6%	5.3%	4.4%	3.2%	3.3%	2.5%	2.0%	1.5%	0.9%	1.8%
Attend spec. schools	0.3%	0.8%	0.5%	0.6%	0.2%	0.1%	0.3%	0.3%	0.3%	0.1%	0.2%	0.5%	0.1%

Notes: Summaries of assignment and enrollment decisions of students in the demand sample under the centralized mechanism. Panel A restricts to students that received an assignment in an NYC Public School in the Main Round. Panel B restricts to students that did not receive an assignment in the main round.

Table A.6: All Estimates from Demand System

		School Char.	School Char.x Demographics	School Char. x Student Achievement	School Characteristics x Demographics School Characteristics x Achievement		
					All Choices	Top Choice	Top Three Choices
		(1)	(2)	(3)	(4)	(5)	(6)
A. Preference for Distance							
Distance		-0.318	-0.313	-0.316	-0.310	-0.387	-0.353
B. Preference for Schools							
School Characteristic	*Student Characteristic*						
9th Grade Enrollment	Uninteracted	1.45E-04	4.01E-04	1.54E-04	5.89E-04	0.001	9.85E-04
	Female				-7.03E-05	-8.86E-05	-8.28E-05
	Asian		1.90E-04		1.55E-04	5.19E-04	3.44E-04
	Black		-3.99E-04		-3.80E-04	-4.23E-04	-4.42E-04
	Hispanic		-1.78E-04		-1.89E-04	-1.57E-04	-2.10E-04
	Standardized Math Score			4.31E-05	1.07E-05	2.15E-05	7.62E-06
	Standardized English Score			-5.50E-05	-2.99E-05	-8.92E-05	-5.21E-05
	No Math Score			-2.69E-04	-1.04E-04	-3.34E-04	-1.94E-04
	No English Score			3.51E-04	8.29E-05	2.67E-04	1.53E-04
	Free Lunch				-1.37E-04	-2.94E-04	-2.40E-04
	Special Ed				-7.05E-05	-1.66E-04	-1.22E-04
	Median Family Income				-6.50E-06	-1.93E-05	-1.93E-05
	Limited English Proficiency				7.43E-05	1.75E-04	1.56E-04
Percent White	Uninteracted	0.026	0.040	0.025	0.032	0.040	0.033
	Female				-6.70E-04	-9.08E-04	-1.95E-04
	Asian		-0.016		-0.015	-0.027	-0.018
	Black		-0.022		-0.020	-0.035	-0.024
	Hispanic		-0.011		-0.009	-0.016	-0.010
	Standardized Math Score			0.001	3.54E-04	4.04E-04	4.63E-04
	Standardized English Score			0.002	0.002	0.005	0.003
	No Math Score			-0.002	-3.79E-04	-3.00E-03	-6.27E-04
	No English Score			0.004	3.00E-04	0.004	4.79E-04
	Free Lunch				0.001	0.003	0.002
	Special Ed				6.46E-04	0.006	0.003
	Median Family Income				9.11E-04	0.002	0.001
	Limited English Proficiency				-2.09E-04	2.37E-04	8.68E-04

Table A.6: All Estimates from Demand System (Cont'd)

		School Char.	School Char.x Demographics	School Char. x Student Achievement	School Characteristics x Demographics School Characteristics x Achievement		
					All Choices	Top Choice	Top Three Choices
		(1)	(2)	(3)	(4)	(5)	(6)
Attendance Rate	Uninteracted	0.055	0.103	0.062	0.075	0.131	0.117
	Female				0.002	0.009	0.005
	Asian		0.013		0.016	0.076	0.039
	Black		-0.053		-0.036	-0.040	-0.046
	Hispanic		-0.053		-0.037	-0.040	-0.047
	Standardized Math Score			0.012	0.010	0.017	0.015
	Standardized English Score			0.014	0.014	0.020	0.021
	No Math Score			-0.024	-0.016	-0.046	-0.026
	No English Score			0.010	-0.004	0.009*	-4.36E-04
	Free Lunch				-3.48E-04	-0.008	-0.005
	Special Ed				0.004	-0.008*	0.005*
	Median Family Income				0.006	0.008	0.007
	Limited English Proficiency				0.006	0.014	0.012
Precent Free Lunch	Uninteracted	-0.002	-0.007	-0.003	-0.004	0.005	-0.003
	Female				-9.47E-04	2.70E-04	-1.62E-05
	Asian		3.96E-04		-8.29E-04	-0.004	-8.80E-04
	Black		0.002		0.001	-0.008	-0.002
	Hispanic		0.010		0.009	0.007	0.010
	Standardized Math Score			-7.32E-04	-5.94E-04	-3.23E-04	-5.84E-04
	Standardized English Score			-0.001	-6.56E-04	-8.88E-04*	-6.48E-04
	No Math Score			-7.20E-04	0.001	-4.11E-04	2.00E-03
	No English Score			0.005	0.002	0.005	0.003
	Free Lunch				1.49E-04	-0.002	-0.001
	Special Ed				0.001	-9.35E-04	7.95E-04
	Median Family Income				-7.22E-04	-0.001	-7.36E-04
	Limited English Proficiency				-4.37E-04	0.003	0.002

Table A.6: All Estimates from Demand System (Cont'd)

		School Char.	School Char.x Demographics	School Char. x Student Achievement	School Characteristics x Demographics School Characteristics x Achievement		
					All Choices	Top Choice	Top Three Choices
		(1)	(2)	(3)	(4)	(5)	(6)
Math Performance of Graduates	Uninteracted	0.008	0.007	0.004	2.68E-04	7.65E-04	-8.66E-04
	Female				0.002	0.002	0.003
	Asian		0.009		0.010	0.014	0.013
	Black		-0.004		3.35E-04	0.005	0.002
	Hispanic		-7.55E-04		0.004	0.009	0.007
	Standardized Math Score			0.008	0.007	0.012	0.009
	Standardized English Score			0.005	0.005	0.006	0.005
	No Math Score			2.17E-04	1.00E-03	0.008	0.002
	No English Score			0.002	1.90E-04	-0.005	-1.42E-04
	Free Lunch				-0.003	-0.005	-0.005
	Special Ed				-0.003	-1.07E-04	-1.00E-03
	Median Family Income				1.19E-04	-7.12E-04	-2.42E-04
	Limited English Proficiency				-0.002	-0.003	-0.005
Spanish Language Pgm.	Limited English Proficiency				5.147	5.755	5.472
	LEP Hispanic Student				-3.142	-3.817	-3.147
Asian Language Pgm.	Limited English Proficiency				3.687	4.253	4.203
	LEP Asian Student				-2.131	-3.299	-2.717
Other Language Pgm.	Limited English Proficiency				2.289	2.947	2.870
Program Type Dummies		X	X	X	X	X	X
Pgm. Specialty Dummies		X	X	X	X	X	X
C. Model Fit							
Log-Likelihood		-2.73E+06	-2.71E+06	-2.71E+06	-2.69E+06	-3.02E+05	-9.07E+05
#parameters		248	272	280	349	349	349.00
Pseudo R^2		0.85	0.85	0.89	0.89	0.89	0.86
McFadden R^2		0.19	0.19	0.19	0.19	0.19	0.19
R-squared of 2nd-Stage Delta Regression		0.43	0.42	0.68	0.66	0.42	0.45
Estimation Time (in hrs)		3.38	4.51	5.60	9.49	1.59	3.46

Notes: Parameter estimates corresponding to Table 1.6.

Appendix B

Appendices to Chapter 2

B.1 Estimation and Inference

B.1.1 Moments

For simplicity of exposition, I consider the case of only one market and treat all characteristics as observed and exogenous. This treatment replaces ν_{jt} with $\hat{\nu}_{jt}$. Error is estimating ν_{jt} is dealt with in a bootstrap when computing standard errors. I use x_i and z_j to denote resident and program characteristics respectively. I assume that covariates that depend on both the residents and the programs can be written as a known function of x_i and z_j. This function is subsumed in the notation.

Given these characteristics and a parameter vector θ, let $F_{X,\varepsilon,Z,\eta}\left(\theta|F_X,F_Z\right)$ denote the stable match distribution given the marginal distributions of observed characteristics of agents on each side of the market. Throughout, I omit conditioning on the marginal distributions to write the match distribution predicted by θ as $F_{X,\varepsilon,Z,\eta}\left(\theta\right)$. I write the match distribution $F_{X,\varepsilon,Z,\eta}\left(\theta_0\right)$ at the true parameter and the population distribution of characteristics as $F_{X,\varepsilon,Z,\eta}$. Expectations with respect to $F_{X,\varepsilon,Z,\eta}\left(\theta\right)$ are denoted E_θ and with respect to $F_{X,\varepsilon,Z,\eta}\left(\theta_0\right)$ denoted E_0. I denote population moments as a function of θ with $m\left(\theta\right)$, sample analogs with $\hat{m}$ and simulation analogs with $\hat{m}\left(\theta\right)$.

I denote the observed match with a function $\mu : \{1,\ldots,N\} \rightarrow \{1,\ldots,J\}$ and a simulated match function at θ with μ_s^θ. Also, let $\tilde{\mu} = \mu^{-1} \circ \mu : \{1,\ldots,N\} \rightarrow 2^{\{1,\ldots,N\}}$ be a map from i to the set of peers of i (possibly empty since it does not include i).

The three sets of moments discussed in Section 2.6.2 have the following mathematical expressions.

1. Moments of the match distribution of observable characteristics of residents and programs. If X and Z are scalar random variables, we can write the second moment of this distribution as

$$\begin{aligned} m_{ov}(\theta) &= E_\theta[XZ] \\ &= \int X\ Z\ dF_{X,Z}(\theta) \\ \hat{m}_{ov} - \hat{m}_{ov}^S(\theta) &= \frac{1}{N}\sum x_i z_j \left[1\{\mu(i) = j\} - \frac{1}{S}\sum 1\left\{\mu_s^\theta(i) = j\right\}\right]. \end{aligned}$$

   In general, an arbitrary function of $\psi(x, z)$ can be used in place of the product of X and Z. One may also use a variable that varies by resident and program, such as an indicator for whether a program is located in the same state as the resident's state of birth.

   For estimation, I include pair of covariances between the set of observed program and resident characteristics that are included in the specifications. I also include moments for the same birth state and the same medical school state. Further, the covariance between the square of the characteristics of the program on which I include random coefficients and resident characteristics are included.

2. The within program variance of resident observables. Note that $F_{X|Z,\eta}(\theta)$ is the distribution of characteristics X matched with hospitals with the same value of Z, η. In a finite sample, this is a unique hospital with probability 1. For a scalar X, let

$$V_\theta(X|z,\eta) = \int (X - E_\theta(X|z,\eta))^2\, dF_{X|z,\eta}(\theta)$$

   denote the average squared deviation of X within program z, η. The moment based on the within program variation is

$$\begin{aligned} m_w(\theta) &= E_\theta\left[V_\theta(X|z,\eta)\right] \\ &= \int V_\theta(X|z,\eta)\, dF_{Z,\eta} \\ \hat{m}_w &= \frac{1}{N}\sum_i \left(x_i - \frac{1}{|\tilde{\mu}(i)|}\sum_{i'\in\tilde{\mu}(i)} x_{i'}\right)^2 \\ \hat{m}_w^S(\theta) &= \frac{1}{NS}\sum_{i,s}\left(x_i - \frac{1}{|\tilde{\mu}_s^\theta(i)|}\sum_{i'\in\left|\tilde{\mu}_s^\theta(i)\right|} x_{i'}\right)^2. \end{aligned}$$

When X is vector valued, one could stack components, or replace the conditional variance $V_\theta(X|z,\eta)$ with a covariance. I use the within program variance for all characteristics included in the specifications. We may replace X with a function $\phi(X)$.

3. Covariance between resident characteristics and the average characteristics of a resident's peers. If $X = (X_1, X_2)$ where X_1 and X_2 are scalars, the quantity

$$E_\theta\left[X_1 E_\theta\left[X_2|z,\eta\right]\right] = \int X_1 E_\theta\left[X_2|Z,\eta\right] dF_{X,z,\eta}(\theta)$$

is the covariance between a resident's characteristic X_1 and the average characteristics of the resident's peers X_2. The moment can be written as

$$\begin{aligned} m_p(\theta) &= E_\theta\left[X_1 E_\theta\left[X_2|z,\eta\right]\right] \\ \hat{m}_p - \hat{m}_p^S(\theta) &= \frac{1}{N}\sum x_{1,i}\left[\frac{1}{|\tilde{\mu}(i)\setminus\{i\}|}\sum_{i'\in\tilde{\mu}(i)\setminus\{i\}} x_{2,i'} - \frac{1}{S}\sum_s \sum_{i'\in\tilde{\mu}_s^\theta(i)\setminus\{i\}} \frac{1}{|\tilde{\mu}_s^\theta(i)\setminus\{i\}|} x_{2,i'}\right]. \end{aligned}$$

In general, one could consider two separate functions of X instead of X_1 and X_2 or the same variable X. I use the covariance between the continuous characteristics of the residents and peer averages of each characteristic included in the specifications.

Alternatively, one could combine moments of the second and third type using the notation to specify the second type of moments. One would match the entries in the upper triangular portion of within program covariance matrix.

B.1.2 A Bootstrap

The number of programs in a given market is denoted J_t. Each program has a capacity c_{jt} that is drawn iid from a distribution F_c with support on the natural numbers less than $\bar{c}$. The total number of positions in market t is the random variable $C_t = \sum c_{jt}$. In each market, the number of residents N_t is drawn from a binomial distribution $B\left(C_t, p_t\right)$ for $p_t \leq 1$. The vector of resident and program characteristics $(z_{jt}, z_{ijt}, x_i, r_{jt}, \varepsilon_i, \beta_i, \eta_{jt}, \zeta_{jt})$ are independently sampled from a population distribution. The distribution of program observable characteristics (z_{jt}, z_{ijt}) may depend on c_{jt} while all other characteristics are drawn independently.

Agarwal and Diamond (2013) study asymptotic theory under this sampling process in the case of a single market $J \to \infty$. Limit theorems for the estimator is not yet complete. Monte Carlo simulations based on inference procedures for standard simulation estimators for the model with exogenous characteristics and preference heterogeneity have a decreasing root mean square error with increase in sample size. In these simulations, I used a parametric bootstrap that accounts for the dependent data structure to estimate the asymptotic variance of the moments, and a delta method to estimate the asymptotic variance of the parameter.

The data can be seen as generated from an equilibrium map from θ and the distribution market participants. Standard Donsker theorems apply for the sampling process for market participants. The inference method above should then be consistent if a functional delta method applies to this map i.e. the distribution of the observed matches is (Hadamard) differentiable jointly in the parameter θ and the distribution of observed characteristics of market participants (at the population distribution of characteristics, tangentially to the space of regular models). Monte Carlo evidence is consistent with this.

I approximate the limit distribution of $\hat{\theta}_{msm}$ as the number of programs in each market grows using

$$
\begin{aligned}
\sqrt{J}\left(\hat{\theta}_{msm}-\theta_0\right) &\approx \left[\left(\Gamma' W \Gamma\right)^{-1}\Gamma' W\right]\sqrt{J}\left(\hat{m}\left(\hat{\theta}_{msm}\right)-m\left(\theta_0\right)\right)\\
&\xrightarrow{d} N\left(0,\Sigma\right)\\
\Sigma &= \left(\Gamma' W \Gamma\right)^{-1}\Gamma' W V^{tot} W' \Gamma \left(\Gamma' W \Gamma\right)^{-1}\\
V^{tot} &= V+\frac{1}{S}V^{S}
\end{aligned}
\tag{B.1}
$$

where W is the weight matrix used in the objective function, $\Gamma = \Gamma\left(\theta_0\right)$ is the gradient of $m\left(\theta\right)$ evaluated at θ_0, and V^{tot} is the asymptotic variance in $\hat{m}^S\left(\theta_0\right)$, and $J = \sum J_t$. The asymptotic variance V^{tot} in $\hat{m}\left(\theta_0\right)$ is the sum of the variance due to two independent process: the sampling variance V arising from sampling the observable characteristics of residents and programs in the economy and the simulation variance V_S due to the sampling unobservable traits of the residents and programs. Note that the sampling variance needs to include the variance in $\hat{m}$ arising from uncertainty in estimating $\hat{\nu}_{jt}$ in different observed samples of programs. The simulation variance is scaled down by S, the number of simulations used to compute $\hat{m}^S\left(\theta\right)$ during estimation. Since closed form solutions for the moments are not available, I use numerical and simulation techniques to calculate each of the unknown quantities Γ, V_S, V^{tot}.

To estimate $\Gamma\left(\theta_0\right)$, I construct two-sided numerical derivatives of the simulated moment function $\hat{m}\left(\theta\right)$ using the observed population of residents and programs. Since $\hat{m}^S\left(\theta\right)$ is not smooth due to simulation errors, extremely small step sizes and a low number of simulation draws can lead to inaccuracies. For this step, I use 10,000 simulation draws and a step size of 10^{-3}. The simulation variance is estimated by calculating the variance in 10,000 evaluations of $\hat{m}^S\left(\hat{\theta}_{msm}\right)$, each with a single simulation draw and using the observed sample of resident and program characteristics. Since these two calculations keep the set of observed residents and programs constant, these two quantities can be calculated independently in each of the markets.

As noted, the sampling variance in $\hat{m}\left(\theta\right)$ needs to account for the fact that the control variable $\hat{\nu}_{jt}$ is estimated. It also needs to account for the dependent structure of the match data. I use the following bootstrap procedure to estimate V.

1. For each market t, sample J_t program observable characteristics from the observed

data $\{z_{jt}, r_{jt}, q_{jt}\}_{j=1}^{J_t}$ with replacement. Denote this sample with $\left\{z_{jt}^b, r_{jt}^b, q_{jt}^b\right\}_{j=1}^{J_t}$

(a) Calculate $\left(\hat{\gamma}^b, \hat{\tau}^b\right)$ and the estimated control variables $\hat{\nu}_{jt}^b$ as in the estimation step.

2. Draw N_t^b from $B\left(\sum_{j=1}^{J_t} q_{jt}^b, \frac{N_t}{Q_t}\right)$ and a sample of resident and resident-program specific observables $\left\{x_{it}^b, \left\{z_{ijt}^b\right\}_{j=1}^{J_t}\right\}_{i=1}^{N_t^b}$ from the observed data, with replacement.

3. Simulate the unobservables to compute $\left\{\hat{m}^{1,b}\left(\hat{\theta}_{msm}\right)\right\}_{b=1}^{B}$ the vector of simulated moments using the bootstrap sample economy. The variance of these moments is the estimate I use for V.

Essentially, the bootstrap mimics the data generating process to sample a new set of agents from the population distribution to form an economy. It replaces the set of observed characteristics of the residents and programs with the empirical distribution observed in the data. Given this economy, it computes $\hat{\nu}_{jt}$ and the moments at a pairwise stable match at $\hat{\theta}$. The covariance of the moments across bootstrap iterations is the estimate of $\hat{V}$. The uncertainty due to simulation error $\hat{V}^S$ is approximated by drawing just the unobserved characteristics.[1] In a large economy, consistency of each of these quantities implies the consistency of the estimate

$$\hat{\Sigma} = \left(\hat{\Gamma}'W\hat{\Gamma}\right)^{-1}\hat{\Gamma}'W\left(\hat{V} + \frac{1}{S}\hat{V}^S\right)W'\hat{\Gamma}\left(\hat{\Gamma}'W\hat{\Gamma}\right)^{-1}. \tag{B.2}$$

Weight Matrix

It is well known that the choice of weight matrix can affect efficiency. This choice is particularly important when the number of moments is much larger than the number of parameters. A common method uses a first stage consistent estimate of θ_0 to obtain variance estimates $\hat{V}$ and $\hat{V}^S$ to compute the optimal weight matrix $\hat{W} = \left(\hat{V} + \frac{1}{S}\hat{V}^S\right)^{-1}$ that can be used in the second stage. One may implement the first step of obtaining a consistent

[1] Justifying the use of a finite number of simulation draws S as $J \to \infty$ needs a stochastic equicontinuity condition on the empirical objective function (see Pakes and Pollard, 1989). Given the incomplete econometric theory, I use 1,000 simulations to mitigate concerns on this front.

estimate of θ_0 using any positive definite matrix W, with the identity matrix as the most commonly used first-step weight matrix. In this application, a two-step procedure is computationally prohibitive. In Monte Carlo simulations with this dataset, I found that using the identity matrix was often inaccurate and left us with a poor estimate of θ_0. Instead, a weight matrix $\tilde{W}$ calculated using the following bootstrap procedure seemed to approximate the optimal weights fairly well. For each market t, with replacement, randomly sample J_t programs and the residents matched with them. Treat the observed matches as the matches in the bootstrap sample as well.[2] Compute moments $\left\{\tilde{m}^b\right\}_{b=1}^{B}$ from the sample and compute the variance $\tilde{V}$ and set $\tilde{W} = \tilde{V}^{-1}$. While this weight matrix need not converge to the optimal weight matrix, the only theoretical loss is in the efficiency of the estimator. This weight matrix also turns out to be close to one that would be calculated as $\hat{W} = \left(\tilde{V}\left(\hat{\theta}_{msm}\right) + \frac{1}{S}\hat{V}^S\left(\hat{\theta}_{msm}\right)\right)^{-1}$ where $\hat{\theta}_{msm}$ is the estimate of θ_0 using $\hat{W}^{sub}$ as the weight matrix, and $\tilde{V}\left(\hat{\theta}_{msm}\right)$ and $\hat{V}^S\left(\hat{\theta}_{msm}\right)$ are the sample and simulation variance that are estimated as described earlier.

B.1.3 Optimization Algorithm

The function defined in equation (2.10) may be non-convex and may have local minima. Further, since $\hat{m}^S(\theta)$ is not smooth as it is simulated. Gradient based global search methods can perform very poorly in such settings. I use an extensive derivative free global search followed by a refinement step that uses a derivative free local search to compute the estimate $\hat{\theta}_{msm}$.

The global search is implemented using MATLAB's genetic algorithm and a bounded parameter space based on initial runs (Goldberg, 1989). The algorithm is derivative free, making it particularly useful for non-smooth problems. Further, the stochastic search method retains parameter values with low fitness (poor values of the objective function) for a significant number of generations in the population but explores the rest of the parameter space using random innovations. This feature makes it attractive for use in settings where local optima may cause some other algorithms to "get stuck" in these local minima.

[2]Note that a submatch of a stable match is also stable. Hence, the constructed bootstrap match is also stable.

In Monte Carlo experiments the algorithm seemed to out-perform other commonly used global optimization techniques such as multi-start algorithms with local search, directed search and simulated annealing.

As with the vast majority of optimizers working with non-convex problems, there is no guarantee that the genetic algorithm finds the global optimum. I conducted three initial genetic algorithm runs to with separately seeded populations of size 40, cross-over fraction of 0.75, one elite child, an adaptive mutation scale of 4 and shrinkage of 0.25. These extensive runs were used to generate starting values for the local searches.

Local searches using starting values yielding the lowest two to three objective function and from similar models were implemented. The step is conducted to refine the estimate $\hat{\theta}_{msm}$ and to be thorough in the search for the global minimum. I used the subplex algorithm (Rowan, 1990), a derivative free optimization routine. It is a variant of the Nelder-Meade algorithm that is more robust for problems with more than a few dimensions. The refined parameter was always close to the one found by the global optimization routine. However, it may be liable to not converge to a minimum. For this reason, I use up to three successive runs of the subplex algorithm implemented in the toolbox NLOpt for these local runs (Johnson, 2011). Each run restarts the algorithm using the optimum found in the previous run. I do not repeat the local search if the change the point estimate between the starting value and the optimum is less than 10^{-6} in Euclidean norm. Two iterations were always sufficient. I also verified that the reported point is at least a local minimum using one dimensional slices of the parameter space and profiling the objective function in the direction of other global search results and local minima that may have been found.

My experience with Monte Carlo experiments suggests that this method is very successful in finding a parameter value close to the true parameter. Although I did not extensively benchmark this procedure against other optimization procedures, the method also seems faster than grid search, multi-start with a local optimization using subplex and the simulated annealing algorithm.

B.2 Parameter Estimates

Table B.1 presents point estimates of the models discussed in Section 2.7 and three additional models. Two of the additional models do not allow for heterogeneity in preferences. The final additional model is a version of specification (1) in Table 2.7 that uses the instrument.

Panel A presents parameter estimates for the distribution of residents' preferences and Panel B presents estimates for the human capital index. As mentioned in the text, these point estimates are not directly interpretable in economically meaningful terms. Table 2.7 translates a subset of coefficients from Panel A into monetized values by dividing a given coefficient by the coefficient on salaries, and scaling them into dollar equivalents for a one standard deviation change.

First, comparing coefficients on salaries from specifications (1) through (3) to the corresponding specifications (4) through (6), we see that accounting for endogeneity in salaries reduces the point estimate on the salary coefficient. Many of the other coefficients are not substantially altered by the inclusion of the control variable and the program's own reimbursement rates. The annual rent and NIH funding of major affiliates are two exceptions. This may be a consequence of correlation between reimbursement rates and these covariates.

Unfortunately, the estimates from specification (6) are not economically interpretable because of the negative coefficient on salaries but is consistent with the general drop in coefficient when using wage instruments. The primary economic implication of the drop in coefficient in salaries on including the instrument, at least for specifications (4) and (5), is that the willingness to pay for programs increases substantially. Specification (4) results in willingness to pay measures that are implausibly large. I attribute this non-robustness to a weak instrument due to the limited variation in salaries. Methods for weak-identification robust estimation are not well developed for non-linear models such as this and are computationally burdensome (Stock et al., 2002).

Comparing estimates from specifications (1) and (2), we see changes in the estimated coefficient on NIH funding of major affiliates, salaries and the medicare wage index, and rent. Note that the change in coefficient on rent does not appear to have economically

Table B.1: Detailed Preference Estimates

	w/o Wage Instruments			w/ Wage Instruments		
	Full Het.	Geo. Het.	No Het.	Full Het.	Geo. Het.	No Het.
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Preference for Programs						
First Year Salary ($10,000)	2.3099	4.5888	0.6180	0.4983	1.9531	-1.1157
	(0.3205)	(0.4500)	(0.0593)	(0.3174)	(0.3533)	(0.1338)
Log Beds (Primary Inst)	2.5652	2.6058	-0.4044	1.6392	2.7780	-0.2000
	(0.3371)	(0.2213)	(0.0512)	(0.2656)	(0.2399)	(0.0534)
Log NIH Fund (Major)	0.0876	2.3046	0.3729	-0.0474	0.6645	0.5228
	(0.1284)	(0.1646)	(0.0257)	(0.1350)	(0.0735)	(0.0343)
Log NIH Fund (Minor)	1.0351	2.2898	0.4160	1.3589	1.3357	0.5428
	(0.1272)	(0.1410)	(0.0274)	(0.1461)	(0.1447)	(0.0315)
Medicare Case Mix Index	4.9815	4.7917	2.4396	7.9283	5.3517	3.1541
	(0.6724)	(0.5733)	(0.1409)	(0.9053)	(0.5163)	(0.1961)
Medicare Wage Index	-5.5213	1.9601	-0.2240	-5.1235	1.4322	-1.1891
	(1.0418)	(0.5107)	(0.1385)	(0.9917)	(0.3742)	(0.1456)
Annual Median Rent ($10,000)	5.9901	-0.5741	1.8420	7.1745	6.1311	3.0188
	(0.8155)	(0.3137)	(0.1371)	(0.7448)	(0.6117)	(0.1946)
Rural Program	1.6925	2.5747	0.2365	1.2727	3.3816	0.7187
	(0.3457)	(0.3540)	(0.0804)	(0.3573)	(0.4332)	(0.0952)
University Based Program	3.6464	5.0845	0.7694	3.6610	4.9082	1.0441
	(0.4098)	(0.5451)	(0.1022)	(0.4372)	(0.5636)	(0.1067)
Community/University Program	-1.1552	-1.0174	-0.3486	-1.7033	-1.4662	-0.5667
	(0.1969)	(0.1645)	(0.0480)	(0.2180)	(0.2114)	(0.0631)
Reimbursement Rate				-0.0966	0.2569	0.1138
				(0.0466)	(0.0433)	(0.0142)
Control Variable				2.4889	8.7394	2.1200
				(0.5335)	(0.7762)	(0.1571)
Rural Progam x Rural Resident	0.2746	0.0500		0.2484	0.0455	
	(0.0476)	(0.0113)		(0.0506)	(0.0093)	
Program in Medical School State	2.2682	1.0563		2.2592	0.8846	
	(0.1869)	(0.0747)		(0.1950)	(0.0555)	
Program in Birth State	1.4650	0.6057		1.4643	0.4787	
	(0.1250)	(0.0443)		(0.1269)	(0.0296)	
Sigma Log NIH Fund (Major)	0.9814			1.1229		
	(0.1833)			(0.1928)		
Sigma Log Beds	4.1294			3.8453		
	(0.5608)			(0.5114)		
Sigma Medicare Case Mix	4.6807			3.2150		
	(0.9656)			(0.9127)		

Table B.1: Detailed Preference Estimates (cont'd)

	w/o Wage Instruments			w/ Wage Instruments		
	Full Het.	Geo. Het.	No Het.	Full Het.	Geo. Het.	No Het.
	(1)	(2)	(3)	(4)	(5)	(6)
Panel B: Human Capital						
Log NIH Fund (MD)	0.1153	0.1269	0.1468	0.1191	0.0941	0.1429
	(0.0164)	(0.0139)	(0.0116)	(0.0156)	(0.0131)	(0.0129)
Median MCAT (MD)	0.0814	0.0666	0.0697	0.0797	0.0413	0.0718
	(0.0070)	(0.0038)	(0.0027)	(0.0056)	(0.0030)	(0.0030)
US Born (Foreign Grad)	0.1503	-0.2470	0.4651	0.2083	0.2927	0.5964
	(0.1021)	(0.0801)	(0.0458)	(0.0989)	(0.0705)	(0.0486)
Sigma (DO)	0.8845	0.7944	0.7454	0.9321	0.7275	0.8168
	(0.0359)	(0.0285)	(0.0319)	(0.0370)	(0.0292)	(0.0399)
Sigma (Foreign)	3.6190	3.0709	1.2850	3.5549	2.8215	1.5483
	(0.1469)	(0.1102)	(0.0550)	(0.1411)	(0.1131)	(0.0756)
Medical School Type Dummies	Y	Y	Y	Y	Y	Y
Moments	106	106	106	118	118	118
Parameters	25	22	19	27	24	21
Objective Function	951.31	1122.78	6136.30	1032.24	1090.10	6191.08

Notes: See Table 2.7 for Panel A estimates monetized in dollar units. Indicator for zero NIH funding of major associates and for minor associates. In uninstrumented specifications, the variance of the vertical unobservable ξ_{jt} is normalized to 1 and in instrumented specifications, the variance of ζ_{jt} is normalized to 1. In all specifications, the variance of unobservable determinants of the human capital index of MD graduates is normalized to 1. All specifications normalize the mean utility from a program with zeros on all characteristics to 0. All specifications normalize the mean human capital index of residents with zeros for all characteristics to 0. Point estimates using 1000 simulation draws. Standard errors in parenthesis. Optimization and estimation details described in an appendix.

meaningful impact on the willingness to pay for programs located in high rent areas as compared to programs in low rent areas. Table 2.9 shows that specifications (1) and (2) yield similar quantities on this front. A reason for this is that medicare wage index and rents are highly correlated with each other. We also see that the relative magnitude on coefficients on rural birth interacted with rural program, program location in birth state and program location in medical school state have similar relative magnitudes although large in overall magnitude in specification (1). I attribute this difference to additional unobserved heterogeneity in specification (1), due to which similar geographic sorting needs to be explained with higher preference for these characteristics.

Table B.2: Out-of Sample Fit: Regressions

	MD Degree			Foreign Degree		
	(1)			(2)		
	Data	Simulated	(s.e.)	Data	Simulated	(s.e.)
First Year Salary ($10,000)	0.129	0.110	(0.036)	-0.178	-0.094	(0.038)
Median Annual Rent	0.261	0.359	(0.074)	-0.328	-0.355	(0.076)
Log # Beds	-0.017	0.084	(0.021)	0.009	-0.083	(0.022)
Log NIH Fund (Major)	0.050	0.047	(0.012)	-0.042	-0.051	(0.013)
Log NIH Fund (Minor)	0.046	0.022	(0.017)	-0.051	-0.022	(0.017)
Rural Program	-0.019	0.128	(0.042)	-0.004	-0.110	(0.044)
Case Mix Index	0.238	0.211	(0.056)	-0.220	-0.205	(0.058)
Medicare Wage Index	-0.233	-0.365	(0.116)	0.257	0.387	(0.124)
	Log NIH Fund (MD)			Median MCAT Score		
	(3)			(4)		
	Data	Simulated	(s.e.)	Data	Simulated	(s.e.)
First Year Salary ($10,000)	0.135	0.123	(0.096)	0.512	0.484	(0.196)
Median Annual Rent	-0.438	0.206	(0.224)	0.065	0.849	(0.421)
Log # Beds	-0.067	0.084	(0.065)	0.130	0.180	(0.128)
Log NIH Fund (Major)	0.397	0.143	(0.040)	0.518	0.172	(0.074)
Log NIH Fund (Minor)	0.097	0.198	(0.042)	0.137	0.147	(0.085)
Rural Program	-0.172	0.225	(0.122)	-0.224	0.065	(0.242)
Case Mix Index	0.237	0.458	(0.179)	-0.218	0.533	(0.340)
Medicare Wage Index	1.225	0.309	(0.342)	3.060	1.145	(0.678)

Notes: Linear Regressions using 2011-2012 data. Each simulation draws a parameter from the estimated asymptotic distribution of specification (1), and unobservables independently. The vector of coefficients is computed for each draws. The table reports the mean estimate and bootstrapped standard error of simulated estimates in parenthesis.

B.3 Wage Competition

B.3.1 Expressions for Competitive Outcomes

I first characterize the competitive equilibria of the model. The expression in equation (2.17) follows as a corollary. For clarity, I refer to the quality of program 1 as q_1 although I normalize it to 0 in the model presented in the text.

Proposition B.9. *The wage w_k paid to resident k by program k in a competitive equilibrium is characterized by*

$$\begin{aligned} w_1 &\in [-aq_1, f(h_1, q_1)] \\ w_k - w_{k-1} + a(q_k - q_{k-1}) &\in [f(h_k, q_{k-1}) - f(h_{k-1}, q_{k-1}),\ f(h_k, q_k) - f(h_{k-1}, q_k)] \end{aligned}$$

Proof. Since the competitive equilibrium maximizes total surplus, resident i is matched with program i in a competitive equilibrium. The wages are characterized by

$$\begin{aligned} IC(k,i) &: \quad f(h_k, q_k) - w_k \geq f(h_i, q_k) - w_i + a(q_k - q_i) \\ IR(k) &: \quad aq_k + w_k \geq 0,\ w_k \leq f(h_k, q_k). \end{aligned}$$

First, I show that $IR(k)$ is slack for $k > 1$ as long as $IR(1)$ and $IC(k,i)$ are satisfied for all i, k. Since $IC(1,k)$ is satisfied,

$$\begin{aligned} f(h_1, q_1) - w_1 &\geq f(h_k, q_1) - w_k + a(q_1 - q_k) \\ \Rightarrow w_k &\geq w_1 + f(h_k, q_1) - f(h_1, q_1) + a(q_1 - q_k) \\ &\geq -aq_k \end{aligned} \tag{B.3}$$

where the last inequality follows from $f(h_k, q_1) - f(h_1, q_1) \geq 0$ and $w_1 + aq_1 \geq 0$ from the $IR(1)$. Also, $IC(k,1)$ implies that

$$\begin{aligned} f(h_k, q_k) - w_k &\geq f(h_1, q_k) - w_1 + a(q_k - q_1) \\ \Rightarrow w_k &\leq f(h_k, q_k) - f(h_1, q_k) + w_1 - a(q_k - q_1) \\ &\leq f(h_k, q_k) - f(h_1, q_1) + w_1 - a(q_k - q_1) \\ &\leq f(h_k, q_k) \end{aligned} \tag{B.4}$$

where the last two inequalities follow since $w_1 \leq f(h_1, q_1)$ from $IR(1)$ and $-a(q_k - q_1) \leq 0$. Equations (B.3) and (B.4) imply $IR(k)$.

Second, I show that it is sufficient to only consider local incentive constraints, i.e. $IC(i, i-1)$ and $IC(i, i+1)$ for all i imply $IC(k, m)$ for all $k,\ m$. Assume that $IC(i, i-1)$ is satisfied for all i. For firms $i \in \{m, \ldots, k\}$, this hypothesis implies that

$$f(h_i, q_i) - w_i \geq f(h_{i-1}, q_i) - w_{i-1} + a(q_i - q_{i-1}).$$

Summing each side of the inequality from $i = m$ to k yields that

$$f(h_k, q_k) - w_k \geq \sum_{i=m+1}^{k} [f(h_{i-1}, q_i) - f(h_{i-1}, q_{i-1})] + f(h_{m-1}, q_m) + a(q_k - q_{m-1}) - w_{m-1}.$$

Since each $f(h_{i-1}, q_i) - f(h_{i-1}, q_{i-1}) \geq f(h_{m-1}, q_i) - f(h_{m-1}, q_{i-1})$ for $i \geq m$,

$$\begin{aligned}
& f(h_k, q_k) - w_k \\
& \geq \sum_{i=m+1}^{k} [f(h_{m-1}, q_l) - f(h_{m-1}, q_{i-1})] + f(h_{m-1}, q_m) + a(q_k - q_{m-1}) - w_{m-1} \\
& = f(h_{m-1}, q_k) + a(q_k - q_{m-1}) - w_{m-1}. \qquad \text{(B.5)}
\end{aligned}$$

Hence, $IC(k, m)$ is satisfied for all $m \in \{1, \ldots, k\}$. A symmetric argument shows that if $IC(i, i+1)$ is satisfied for all k, then $IC(k, m)$ is satisfied for all $m \in \{k, \ldots, N\}$

To complete the proof, note that local ICs yield the desired upper and lower bounds. □

Corollary B.1. *The worker optimal competitive equilibrium wages are given by*

$$w_k = f(h_1, q_1) - a(q_k - q_1) + \sum_{i=2}^{k} [f(h_i, q_i) - f(h_{i-1}, q_i)]$$

and the firm optimal competitive equilibrium wages are given by

$$w_k = -a(q_k - q_1) + \sum_{i=2}^{k} [f(h_i, q_{i-1}) - f(h_{i-1}, q_{i-1})]$$

B.3.2 Proof of Proposition 2.1

For clarity, I refer to the quality of program 1 as q_1 although I normalize it to 0 in the model presented in the text. As before, I limit attention to production technologies that lead to positive assortative matching between h and q. To focus on the split of the total production, consider two production technologies for which the total output produced by each matched pair is the same for the two technologies. Thus, each N-vector of outputs $y = (y_1, \ldots, y_k)$ defines a family of production functions $\mathcal{F}(y) = \{f : f(h_k, q_k) = y_k\}$ where y_k denotes the output produced by the pair $(h_k,\ q_k)$. The two extremal technologies above in this family are given by $\bar{f}_y(h_k, q_l) = y_k$ and $\underline{f_y}(h_l, q_k) = y_k$ for all $l \in \{1, \ldots, N\}$. Let $w_k^{fo}(f)$ (likewise $w^{wo}(f)$) denote the firm-optimal (worker-optimal) competitive wage under technology f.

I prove a slightly stronger result here as it may be of independent interest. This result shows that the split of surplus in cases other than $\bar{f}$ and $\underline{f}$ are intermediate.

Theorem B.3. *In the worker-optimal (firm-optimal) competitive equilibria, each worker's wage under $f \in \mathcal{F}(y)$ is bounded above by her wage under $\bar{f}_y$ and below by her wage under $\underline{f_y}$.*

Hence, for all $f \in \mathcal{F}(y)$, the set of competitive equilibrium wages of worker k is bounded below by $w_k^{fo}\left(\underline{f_y}\right) = -aq_k$ and above by $w_k^{wo}\left(\bar{f}_y\right) = y_k - aq_k$.

Proof. I only derive the bounds for the worker optimal equilibrium since the calculation for the firm optimal equilibrium is analogous. From the expressions in corollary B.1,

$$\begin{aligned} w_k^{wo}\left(\underline{f_y}\right) &= \underline{f_y}(h_1, q_1) - a(q_k - q_1) \\ &= y_1 - a(q_k - q_1) \end{aligned}$$

since the terms in the summation are identically 0. For any production function, $f \in \mathcal{F}(y)$,

$$\begin{aligned} w_k^{wo}(f) &= f(h_1, q_1) - a(q_k - q_1) + \sum_{i=2}^{k} \left[f(h_i, q_i) - f(h_{i-1}, q_i)\right] \\ &\geq y_1 - a(q_k - q_1) = w_k^{wo}\left(\underline{f_y}\right) \end{aligned}$$

since $f(h_1, q_1) = y_1$ and $f(h_i, q_i) - f(h_{i-1}, q_i) \geq 0$. Similarly, note that

$$w_k^{wo}\left(\bar{f}_y\right) = y_k - a(q_k - q_1)$$

and since each $f(h_i, q_i) - f(h_{i-1}, q_i) \leq f(h_i, q_i) - f(h_{i-1}, q_{i-1})$,

$$\begin{aligned} w_k^{wo}(f) &\leq f(h_k, q_k) - a(q_k - q_1) \\ &= y_k - a(q_k - q_1) = w_k^{wo}\left(\bar{f}_y\right). \end{aligned}$$

$\square$

Proposition 2.1 follows as a corollary.

Proof. For any $y = (y_1, \ldots, y_k)$ and production function $f \in \mathcal{F}(y)$, the profit of firm k is given by

$$\begin{aligned} f(h_k, q_k) - w_k &= y_k - w_k \\ &\geq y_k - w_k^{wo}\left(\bar{f}_y\right) \\ &= a(q_k - q_1) \end{aligned}$$

$\square$

B.3.3 Worker Optimal Equilibrium: Algorithm

The first step uses a linear program to solve for the assignment that produces the maximum total surplus. Let a_{ij} be the total surplus produced by the match of resident i with program j. This surplus is the sum of the value of the product produced by resident i at program j and the dollar value of resident i's utility for program j at a wage of 0.[3] With an abuse of notation of the letter x, let x_{ij} denote the (fraction) of resident i that is matched with program j. Sotomayor (1999) shows that the surplus maximizing (fractional) matching is the solution to the linear program

$$\max_{\{x_{ij}\}} \sum x_{ij} a_{ij} \tag{B.6}$$

subject to

$$0 \leq x_{ij} \leq 1$$

$$\sum_j x_{ij} \leq 1$$

$$\sum_i x_{ij} \leq c_j.$$

[3] As mentioned in footnote 41, I assume that the equilibrium is characterized by full employment. If utilities are normalized so that an allocation is individual rationality if the resident obtains non-negative utility, then α_{ij} at the resident i's least preferred program j must exceed the negative of the dollar monetized utility resident i obtains at j at a wage of zero.

Interpreting x_{ij} as the fraction of total available time resident i spends at program j, the first two constraints are feasibility constraint on the resident's time. The third constraint says that the program does not hire more than its capacity c_j. For a generic value of a_{ij}, the program has an integer solution. This formulation is computationally quicker than solving for the binary program with x_{ij} restricted to the set $\{0, 1\}$. I check to ensure that the solutions I obtain are binary.

The second step seeks to find the worker optimal wages supporting this assignment. The algorithm is based on the dual formulation of the one-to-one assignment problem, which has an economic interpretation given by Shapley and Shubik (1971). Assume for now that $c_j = 1$ for all j. If u_i is the utility imputation for resident i and v_j is the imputation for program j, then a core allocation ensures that for all i, j $u_i + v_j \geq a_{ij}$. This inequality holds for a core allocation if i and j are matched since utility in fully transferable, and if i and j are not matched since otherwise i and j would block the allocation.[4] A particular element in the core can be found by solving the problem

$$\begin{aligned}
&\min_{\{u_i\},\,\{v_j\}} \sum u_i + \sum v_j \\
&\text{subject to} \\
&u_i \geq 0,\ v_j \geq 0 \\
&u_i + v_j \geq a_{ij}
\end{aligned}$$

where the first inequalities are the individual rationality inequalities and the second is the no blocking or incentive compatibility inequality.

In the many-to-one assignment problem I solve, the total production from a set of residents R for a program j is given by $\sum_{i \in R} f_{ij}$ where f_{ij} is the production from i matching with j. Hence, the total surplus from assignments to program j is given by $\sum_{i \in R} a_{ij}$. Since the total surplus at a program is the sum of the surpluses from each residency position, one could rewrite this many-to-one problem as a one-to-one problem between residents and residency positions. This reformulation needs the additional restriction that a resident may not block an allocation with another position at the same program. Let k denote a

[4]See Roth and Sotomayor (1992) for a more detailed discussion of core allocations and the no blocking condition. Sotomayor (1999) constructs the dual formulation of the many-to-one problem.

residency position and j_k denote the program that offers this position. An assignment to positions $\{y_{ik}\}$ with imputations $\{u_i\}$ and $\{v_k\}$ is blocked if there exist i and k such that $u_i + v_k < a_{ij_k}$ and $y_{ik'} = 0$ for all positions k' at program j_k. In other words, an allocation is blocked only by a resident and position pair in which the position is at a program other than the resident's assignment.

Let $\left\{x^*_{ij}\right\}$ denote the optimal assignment assignment found in the first step and $\{y^*_{ik}\}$ be an associated optimal position assignment. The solution to the following linear program gives us imputations corresponding to the worker-optimal allocation:

$$\begin{aligned} &\max_{\{u_i\},\{v_k\}} \sum u_i \\ &\text{subject to} \\ &u_i \geq 0,\ v_k \geq 0 \\ &\sum u_i + \sum v_k \leq \sum x^*_{ij} a_{ij} \\ &u_i + v_k = a_{ij_k} \text{ if } y^*_{ik} = 1 \\ &u_i + v_k \geq a_{ij_k} \text{ if } x^*_{ij_k} = 0. \end{aligned} \tag{B.7}$$

The second constraint is implied by the optimality of the assignment x^* as no feasible imputation may provide a larger total surplus. This constrain always binds since the problem maximizes the surplus that accrues to the residents and none of the other constraints bound this surplus. The third constraint asserts that the imputations supporting y^* result from lossless transfers between a resident her matched program. The final constraints are no blocking constraints between worker i and a position at an unmatched program. Calculating the transfers implied by a solution to this problem is straightforward.

The linear programs were solved using Gurobi Optimizer (http://www.gurobi.com).

B.3.4 Implicit Tuition

I prove a more general result for many-to-one assignment games that subsumes Proposition 2.2. To do this, I first need to introduce some notation. A many to one assignment game between workers $i \in \{1, \ldots, N\}$ and firms $j \in \{1, \ldots, J\}$. The capacity of firm j is c_j. I focus on the case when $\sum_j c_j \geq N$. Worker i, with human capital h_i, produces

$f(h_i) \geq 0$ at firm j, independently of the other workers at the firm. An empty slot produces 0. The utility worker i receives from working at firm j at a wage of w is $u_{ij} + w$. Since the wage transfer is lossless, the total surplus produced by the pair $i,\ j$ under the production function f is $a_{ij}^f = u_{ij} + f(h_i)$. I assume that each $u_{ij} \geq 0$.

Rigorous treatments of these concepts are given in Roth and Sotomayor (1992), but I recall definitions for clarity. For a one-to-one assignment game, an **assignment** is a vector $x = \{x_{ij}\}_{i,j}$ where $x_{ij} = \{0, 1\}$ and $x_{ij} = 1$ denotes that i is assigned to j. The assignment x is **feasible** if $\sum_i x_{ij} \leq 1$ and $\sum_j x_{ij} \leq c_j$. An **allocation** is the pair (x, w) of an assignment x and wages $w = \{w_{ij}\}_{ij}$ with $w_{ij} \in \mathbb{R}$. The allocation is feasible if x is feasible. An **outcome** is a pair $((u, v); x)$ of payoffs $u = \{u_i\}_i$ and $v = \{v_j\}_j$ and an assignment x. Given an allocation, we can compute the outcome $u_i = \sum_j x_{ij}(u_{ij} + w_{ij})$ and $v_j = \sum_i x_{ij}(f(h_i) - w_{ij})$. The outcome is feasible if it can be supported by a feasible allocation (x, w).

In the many-to-one case, we refer to an **assignment of positions** $\{y_{i,p}\}_{i,p}$ where $p \in \left\{1, \ldots, \sum_j c_j\right\}$ denotes a position p and a firm. Let j_p denote the firm offering position p. Each assignment x induces a unique canonical assignment of positions y where the positions in the firm are filled by residents in order of their index i. It's obvious that the function between an assignments and its canonical assignments of positions is bijective. Likewise, with a slight abuse of notation, we can define definition for an allocation of positions using a pair (y, w), where $w = \{w_{ip}\}$. For an allocation (x, w) we can obtain an allocation of positions $(y, \tilde{w})$ by setting y to the canonical assignment and the salaries to $\tilde{w}_{ip} = w_{ij_p}$. The surplus of position p is defined as $v_p^f = \sum y_{ip}(f(h_i) - w_{ip})$ and of worker i by $u_i^f = \sum y_{ip}\left(u_{ij_p} + w_{ip}\right)$. Feasibility of outcomes in this setting can be defined analogously to the previous case. Rigorous treatments of these concepts are given in Camina (2006) and Sotomayor (1999).

A feasible outcome $((u, v); x)$ is **stable** if $u_i \geq 0$, $v_j \geq 0$ and $u_i + v_j \geq a_{ij}$ for all i, j. The allocation (x, w) is a **competitive equilibrium** if the demand of each worker and firm at prices given by w. The equivalence of stable outcomes and competitive equilibria is well known. For the many-to-one case, an with $((u, v); y)$ is stable if for all i, p, $u_i \geq 0$, $v_p \geq 0$, $u_i + v_p \geq a_{ij_p}$ if $y_{ip} = 1$ or $x_{ij_p} = 0$. Consequently, unmatched worker and firms can block if they can produce agree to a mutually beneficial outcome. A matched worker and firm pair can also block an outcome if the sum of their payoffs is lower than the total surplus they produce. The correspondence between many to one stable outcomes and competitive equilibria is noted in Camina (2006). In many to one settings, the demand for firm positions is defined by restricting the wages for each position at a firms to be the same for a given worker. Different workers may, however, face different prices.

Now, we are ready to prove the desired result from which the one-to-one matching case follows trivially by allowing for only one position at each firm.

Proposition B.10. *The equilibrium assignment of positions for the games a_{ij}^{f} and $a_{ij}^{\tilde{f}}$ coincide. Further, if u_i^f and v_p^f are position payoffs for the game a^f, then $u_i^{\tilde{f}} = u_i^f + \left(\tilde{f}(h_i) - f(h_i)\right)$ and $v_p^{\tilde{f}} = v_p^f$ are equilibrium payoffs under the surplus $a_{ij}^{\tilde{f}}$. Consequently the implicit tuition for each position is the same for the games a^f and $a^{\tilde{f}}$.*

Proof. Sotomayor (1999) shows that equilibria for a^f and $a^{\tilde{f}}$ exist and maximize the total surplus in the set of feasible assignments. Towards a contradiction, assume that $y^{\tilde{f}}$ is an equilibrium for $a^{\tilde{f}}$ but not for a^f. The feasibility constraints are identical in the two games, and so both y^f and $y^{\tilde{f}}$ are feasible for both games. Since $y^{\tilde{f}}$ maximizes the total surplus under $a^{\tilde{f}}$,

$$\begin{aligned} \sum_{i,p} a_{ij_p}^{\tilde{f}} y_{ip}^{\tilde{f}} &> \sum_{i,p} a_{ij}^{\tilde{f}} y_{ip}^{f} \\ \Rightarrow \sum_{i,p} a_{ij_p}^{f} y_{ip}^{\tilde{f}} + \sum_i \sum_p \left(\tilde{f}(h_i) - f(h_i)\right) y_{ip}^{\tilde{f}} &> \sum_{i,p} a_{ij_p}^{f} y_{ip}^{f} + \sum_i \sum_p \left(\tilde{f}(h_i) - f(h_i)\right) y_{ip}^{f} \end{aligned} \tag{B.8}$$

Since every worker-firm pair produces positive surplus and the total capacity exceeds the number of workers, there cannot be any unassigned workers in any feasible surplus maximizing allocation, i.e. $\sum_p y_{ip}^f = \sum_p y_{ip}^{\tilde{f}} = 1$ for all i. Hence, we have that $\sum_p \left(\tilde{f}(h_i) - f(h_i)\right) y_{ip}^{\tilde{f}} = \sum_i \left(\tilde{f}(h_i) - f(h_i)\right) y_{ij}^{f}$. The inequality in equation (B.8) reduces to $\sum_{i,p} a_{ij_p}^{f} y_{ip}^{\tilde{f}} > \sum_{i,p} a_{ij_p}^{f} y_{ip}^{f}$, a contradiction to the assumption that y^f is an equilibrium assignment for y^f. This contradiction implies that the equilibrium assignments of positions under the two games coincide.

To show that the second part of the result, consider the payoffs for a^{f^*} where $f^*(h_i) = \max\left\{\tilde{f}(h_i), f(h_i)\right\}$. I show that $u_i^{f^*} = u_i^f + (f^*(h_i) - f(h_i))$ and $v_p^{f^*} = v_p^f$. The comparison of equilibrium payoffs for $\tilde{f}$ and f follows immediately from this. Note that for all i and p, $u_i^f \geq 0$ and $v_j^f \geq 0$ implies $v_j^{f^*} \geq 0$ and $u_i^{f^*} \geq 0$ since $f^*(h_i) - f(h_i) \geq 0$. It remains to that $u_i^{f^*} + v_p^{f^*} \geq a_{ij_p}^{f^*}$ if i is assigned to position p or if i is not assigned to firm j_p. Note that for all i and p, we have that if $u_i^f + v_p^f \geq a_{ip}^f$,

$$\begin{aligned}
u_i^{f^*} + v_p^{f^*} &= u_i^f + f^*(h_i) - f(h_i) + v_p^f \\
&\geq a_{ij_p}^f + f^*(h_i) - f(h_i) \\
&= a_{ij_p}^{f^*}.
\end{aligned}$$

To complete the proof I need to show that the payoffs to each position coincides under the worker-optimal stable outcome. Let u_i^f and v_p^f denote this outcome for the game a^f. Let u_i^0 and v_p^0 be the worker-optimal outcome under the function $f(h_i) = 0$ for all h_i. I showed earlier that the optimal assignments coincide for these two cases. I have shown that $u_i^0 + f(h_i)$ and v_p^0 is stable for a^f. Towards a contradiction, assume that $u_i^f \geq u_i^0 + f(h_i)$ with strict inequality for at least one i. This implies that $u_i^f - f(h_i)$ is stable for a^0. Hence, $u_i^f - f(h_i) \geq u_i^0$ with strict inequality for at least one i, contradicting the assumption that u_i^0 and v_p^0 are part of the worker-optimal outcome. If y is the optimal assignment, this shows that $v_p^0 = \sum_i y_{ip}\left(a_{ip}^0 - u_i^0\right) = \sum_i y_{ip}\left(a_{ip}^f - u_i^f\right) = v_p^f$, proving the result. $\square$

B.4 Rural Hospitals

B.4.1 Suggestive Evidence on Preference Heterogeneity for Rural Doctors

If preferences for resident traits other than a single human capital index were important, one expects that two residents at the same program have dissimilar academic qualifications if they differ on these dimensions. More concretely, one may expect that at rural programs, a rural born resident is academically less qualified than her peer born in an urban location. This may happen because a rural program prefers a rural born resident to an equally qualified urban born resident. To assess whether rural born residents in rural hospitals are more qualified than their urban colleagues, I estimated the regression

$$x_i = \delta\, rural_i + program\_fe_{\mu(i)} + e_i,$$

where x_i is a measure of medical school quality for resident i, $rural_i$ is a dummy for a rural born resident and $program\_fe_{\mu(i)}$ is a fixed effect for program $\mu(i)$, resident i's match.

The results presented in Table B.3 suggests that this may not be of primary importance. Columns (1) and (6) show that rural born residents matched with rural hospitals hail from medical schools that have, on average, only about 0.06 log points less NIH research funding that their peers born in urban areas and are about one percentage point more likely to have an MD degree. Note that the standard deviation in log NIH funding is 1.23. Neither estimate is statistically significant. Although not presented here, the conclusion is robust to using median MCAT score as an indicator of a resident's quality in place of research funding or medical school ranks. If program-year fixed effects are included in place of program-fixed effects, the estimates are more imprecise and the hypothesis that the medical school qualities of the rural born residents at rural hospitals is same as their urban born peers still cannot be rejected. Columns (3) and (8) of the table show this observation is despite the fact that the average rural born residents hails from an observably different medical school than their urban counterparts.

As a validation exercise, I ran similar regressions using gender in place of rural birth.

Table B.3: Preferences for Rural Born Doctors

	Log NIH Funding (MD)					Allopathic/MD Degree		
	Rural Pgms.	Urban Pgms.	All	All	All	Rural Pgms.	Urban Pgms.	All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Rural Born Resident	-0.0582	0.0015	-0.1284***			0.0122	0.0176	0.0263**
	(0.0811)	(0.0364)	(0.0339)			(0.0324)	(0.0119)	(0.0107)
Female				-0.0153	0.0681***			
				(0.0234)	(0.0255)			
Program Fixed Effect	Y	Y		Y		Y	Y	
Observations	750	7,885	8,635	9,599	9,599	1,200	11,260	12,460
R-squared	0.2916	0.2461	0.0017	0.2535	0.0007	0.2568	0.2662	0.0005

Notes: Linear regression of resident's graduating school characteristic on other resident characteristics. Column header Rural (Urban, All) indicates regressions using residents matched to rural (urban, all) programs. Samples pooled from the academic years 2003-2004 to 2010-2011. Columns (1-5) restrict to residents graduating from medical schools with non-zero average annual NIH funding. Columns (1-3) and (6-8) restrict to the subset of residents with reliable city of birth information and were born in the United States. Standard errors clustered at the program level in parenthesis. Significance at 90% (*), 95% (**) and 99% (***) confidence.

Since accreditation guidelines prohibit programs from discriminating on the basis of sex,[5] one may reasonably expect that there is no gender based discrimination by residency programs. Columns (5) and (6) show that although the average female resident hails from medical schools that is better funded than male residents in their cohort, their medical school quality is no different from their male colleagues in their residency program.

While these results are reassuring, they are not definitive on the lack of preference heterogeneity. The somewhat large standard errors and the fact that these observables are proxies for resident quality are the primary reasons for this reserved interpretation. Nonetheless, they suggest that estimates may not suffer from large biases.

B.4.2 General and Partial Equilibrium Effects of Financial Incentives

I consider a partial equilibrium alternative to simulations presented in Section 2.9.1 that may be analytically inexpensive but could, in some situations, perform fairly well. Suppose a policymaker could survey rural residency program directors to determine the impact of incentives for rural training on the residents that choose to train there. For instance, a survey such as the National GME census used in this study could be also solicit a program director's judgement of the number and quality of residents that would match to the program if it unilaterally raised its salary. The responses could be used to predict the impact of the financial incentives studied earlier by simply aggregating the number of positions filled and resident types in rural areas. Such a calculation ignores the influence of a resident who is on the margin between two rural programs and an urban program on the final results. By ignoring the fact that salaries at all rural programs would be increased simultaneously, the calculation acts as if program directors at both rural programs believe that this resident is matched to their program.

The hypothetical benchmark can be simulated using the estimated model by aggregating predicted changes in the matches from the unilateral salary increases at rural hospitals. Panel A of Table B.4 compares results for \$5,000, \$10,000 and \$20,000 increase in salary

[5]The institutional requirements from the Acceditation Council for Graduate Medical Education (ACGME) states that "ACGME-accredited programs must not discriminate with regard to sex, race, age, religion, color, national origin, disability, or any other applicable legally protected status."

Table B.4: General vs. Partial Equilibrium Effects of Price Incentives

	Full Heterogeneity w/o Wage Instruments		
Subsidy Size	\$5,000	\$10,000	\$20,000
	(1)	(2)	(3)
Panel A: Rural Programs			
Total Capacity	—	334	—
Observed # Matches	—	310	—
Baseline Simulated Matches	—	313.33	—
Baseline Prob Rural Match > Urban Match	—	52.76%	—
Δ # Matches (General Equilibrium)	10.23	17.3	20.63
Δ Prob Rural Match > Urban Match (GE)	9.38%	17.70%	31.28%
Δ # Matches (Partial Equilibrium)	10.31	17.59	20.63
Δ Prob Rural Match > Urban Match (PE)	10.22%	19.56%	34.22%
Panel B: Medically Underserved States and Rural Programs (MUA)			
Total Capacity	—	751	—
Observed # Matches	—	720	—
Baseline Simulated Matches	—	721.79	—
Baseline Prob MUA Match > Other Matches	—	53.53%	—
Δ # Matches (General Equilibrium)	14.72	24.7	29.17
Δ Prob MUA Match > Other Matches (GE)	8.73%	16.82%	29.93%
Δ # Matches (Partial Equilibrium)	16.46	25.88	29.17
Δ Prob MUA Match > Other Matches (PE)	9.31%	18.25%	32.70%
Panel C: 1 in 4 Randomly Chosen Programs			
Δ # Matches (General Equilibrium)	21.54	32.23	38.74
Δ # Matches (Partial Equilibrium)	25.45	34.04	39.05
Prob PE Match > GE Match	52.59%	56.43%	67.58%

Notes: Medically underserved states are in the bottom quartile of physician to population ratios or in the top 10 in total area designated as a Health Physician Shortage Area (HPSA). All simulations use 2010 - 2011 sample with 3,148 residents and 3,297 total number of positions. Baseline and counterfactual simulations using 100 draws of structural unobservables. Inter-quartile range in parenthesis. Prob. X > Y is the Wilcoxian statistic: probability that the human capital of the population X is drawn from is greater than that of the population that Y is drawn from.

to rural programs. Comparing the results with those in Panel A, it appears that this simple partial equilibrium analysis would do fairly well at predicting the overall impact of subsidies to rural programs. The impact on resident quality and numbers are only slightly overstated. This observation is because at the estimated parameters, most residents are indifferent between a rural hospital and an urban hospital rather than two rural hospitals, and the number of rural positions is only about a tenth of all positions in the market. This fact is reflected in the distribution graphed in Figure 2.2.

Panel B of Table B.4, compares outcomes for incentives for training in rural programs as well as medically underserved states. The ACA redistributes previous allocated funding to urban programs but currently unused to residency training to (i) rural programs, (ii) states in the bottom quartile of the physician to population ratio and (iii) states in the top 10 in numbers of people living in a Health Physician Shortage Area. I label these states[6] as medically underserved states and compare the partial and general equilibrium impacts of financial incentives. We see that for a \$5,000 incentive, the partial equilibrium analysis predicts an 11% larger impact of subsidies. Notice that for larger subsidies, the difference between the partial and general equilibrium predictions in the change in the number of matches is smaller. For a larger subsidy, the partial equilibrium analysis overstates the change in quality of residents matched at programs in medically underserved states.

Qualitatively similar, but quantitatively larger answers were obtained from a simulation exercise in which I randomly subsidized one-quarter of the residency programs. Panel C presents these results. These simulation experiment shows that the model is capable of capturing potentially important general equilibrium effects of policy interventions. The size of these effects depends on the primitive preferences in the market structure as well as the scope of the intervention.

[6]CMS identified Montana, Idaho, Alaska, Wyoming, Nevada, South Dakota, North Dakota, Mississippi, Florida, Peurto Rico, Indiana, Arizona and Georgia as in the bottom quartile of physicians to population ratio. Lousiana, Mississippi, Peurto Rico, New Mexico, South Dakota, District of Columbia, Montana, North Dakota, Wyoming and Alabama are in the top 10 in numbers of people living in primary care HPSAs. Peurto Rico is exlcuded from this analysis.

Table B.5: Recruitment Into Rural Practice

	Urban Born Resident		Rural Born Resident	
	Urban Program	Rural Program	Urban Program	Rural Program
Percent Practicing in a Rural County	19.52%	50.45%	46.35%	79.19%

Notes: Means of location outcomes for US born residents entering a non-academic practice and with good data on birth city and practice city. Post-graduation plans from graduating resident survey administered to residency program directors in the National GME Census Track. The headers Urban (Rural) Program indicates whether a resident graduated from an urban (rural) program. Results from 5878 resident observations and 2027 program-year observations.

B.5 Data Construction

B.5.1 National GME Census

The American Medical Association (AMA) and the Association of American Medical Colleges (AAMC) jointly conduct an annual National Graduate Medical Education Census (GME Track) of all residency programs accredited by the Accreditation Council for Graduate Medical Education (ACGME). There are two main components of the census: the program survey and the sponsoring institution survey. The program survey, which is completed by the program directors, also gathers information about the residents training at the programs. Fields from the surveys are used to update FRIEDA Online, a publicly accessible database and the AMA physician masterfile. Since 2000, the GME Track has been pre-augmented with data from the Electronic Residency Application Service (ERAS) and the National Residency Matching Program (NRMP).[7] The AMA provided records from the National GME census on all family medicine residency training programs in the Unites States between 2003-2004 and 2010-2011. The 2011-2012 data was provided after the initial empirical analysis was completed.

The data files and identifiers are structured as follows:

1. Program file with program name, characteristics, a unique identifier for the program. This file also contains the identifier for the program's affiliated hospitals.

2. Resident file with resident characteristics, program code, country code and medical school code. Two separate files identify the country and MD granting medical schools by name.

3. Institution file with the institution name, characteristics and a unique identifier.

4. Two bridge files. One delineating the relationships between programs and institutions (usually hospitals) as primary institution, sponsoring institution or clinical affiliate, and the other delineating the relationships between institutions and medical schools as major affiliate, graduate affiliate or limited affiliation.

[7] The details of the data collection procedure are outlined on http://www.ama-assn.org/ama/pub/education-careers/graduate-medical-education/freida-online/about-freida-online/national-gme-census.page.

Sample Construction

The baseline sample is constructed from the set of all family medicine residency programs accredited by the ACGME and first-year residents training at such programs. From this set, I exclude programs in Puerto Rico, military programs and their first-year residents. Less than 20 programs and 123 residents are excluded due to these cuts. I also exclude programs that do not participate in the National Residency Matching Program and the residents matched to these program. These constitute less than 9 programs and 22 residents in each year. Finally, I also exclude the set of programs not offering any first-year positions, and programs that have no reported first-year matches during the entire sample period from the analysis. This final exclusion leads to 21 programs being dropped from the sample in 2003-2004, and less than 5 programs being dropped in the other years.

A detailed breakdown of the annual counts of the sample selection procedure is provided in Table B.6.

Merging GME Track Data

Programs to Clinical Site I wish to identify the primary hospital at which the clinical training of the residents in the programs occur. The AMA data identifies the relationship between programs and sponsoring institutions and hospitals in two ways. The program files records list each program's primary site. The program-institution bridge file records the sponsoring institution, (a second) primary clinical site and other affiliated institutions.

The program-institution bridge has the drawback that the clinical site of the program is not very well reported in the program-institution bridge with at most 94 observations (amongst all ACGME family medicine programs) in any given year whereas the sponsoring institutions are often medical schools or health systems. In order to avoid prioritizing sponsoring institutions or clinical sites from the bridge file, I pick the primary clinical site as reported in the program file as the starting point.

In a large number of cases, the institutition type of the primary institution was a medical school or a health system, not a hospital. Consequently, the hospital institution data for these observations were not available. In the vast majority of these cases, the primary institution, at some point during the sample period was reported as a different site, one that

Table B.6: Sample Construction

Year	03-04	04-05	05-06	06-07	07-08	08-09	09-10	10-11
Panel A: Programs								
Total number of ACGME Programs	475	462	463	460	457	453	451	451
Excluding programs in Peurto Rico	469	458	459	456	453	449	448	448
Excluding military programs	455	444	445	442	440	436	432	434
Excluding programs that do not participate in the NRMP	446	438	443	438	432	432	427	429
Excluding programs that are not offering positions	445	438	441	438	431	432	427	429
Excluding programs with no matches in the sample period	425	433	439	436	427	430	423	428
Panel B: Residents								
Total number of Residents in ACGME programs	3118	3066	3166	3148	3095	3154	3133	3268
Excluding residents matched with Peurto Rico programs	3097	3048	3154	3140	3085	3143	3126	3254
Excluding residents matched with military programs	2995	2945	3041	3026	2996	3051	3009	3160
Excluding residents matched with NRMP non-participants	2976	2925	3035	3021	2974	3040	2996	3148

was a hospital. I checked all cases in which the primary institutition was not a hospital or clinic as identified by an institution type field in the institution file, or had a bed count of zero. When possible, I changed the primary hospital of a program from the listed program according to the following rules:

1. I first checked the program-institution bridge for a listed primary clinical site that was a hospital and changed the primary hospital to that primary clinical site.

2. I looked at the closest year in which the program listed a primary clinical site that is a hospital or clinic and changed it to that hospital or clinic only if the institution was listed as an affiliate or sponsor in that year as well.

The changes affected a total of 285 out of 3441 program-year primary clinical institution relationships in 109 out of 462 programs in the unrestricted sample of all family residency programs between 2003-2004 and 2010-2011. In any given year, no more than 43 programs were affected in any given year.

Finally, 82 program-year observations did not have institution data from the primary sites based on the designation of primary sites above. These programs were solely sponsored by health systems or medical schools, and not primarily associated with a hospital. I imputed the hospital characteristics by taking the mean characteristics of all hospital affiliates for these programs. This imputation populated records in 11 programs in 2003-2004 and 2004-2005 and 10 records in the other years.

Programs with Medical Schools The link between medical schools and programs is provided by the AMA through the program-institution bridge followed by the institution-medical school bridge. The program-institution relationships are categorized into primary clinical sites, sponsors and affiliates. The institution-medical school relationships are categorized as limited, graduate and major.

I use these relationships to define two types of affiliations for programs to medical schools, major and minor. A program has a major affiliation to a medical school if the primary or sponsoring institution has a major affiliation with a medical school. All other relationships are regarded as minor relationships. The relationships between programs and medical schools are imputed for all years between the first and last year of a major (likewise minor) relationship. I used all relationships since 1996 for this imputation and for 2010-2011, I used the relationships in 2009-2010 as well. For the unselected sample of family medicine programs between 2002-2003 and 2010-2011, I imputed relationships for 144 out of 2797 major affiliations and 702 out of 3337 minor affiliations. The mean NIH funding across all major and minor affiliations are used as the variables for this merge.

B.5.2 Medical School Characteristics

The National GME Census does not provide data on medical school characteristics. Each medical school is identified by a number, and only the medical school names for MD granting medical schools are identified. According to the AAMC, there are 134 accredited MD-granting medical schools in the United States. In the dataset, I found 135 medical school identifiers for MD granting institutions. Texas Tech University Health Sciences Center School of Medicine appeared with two different ids. I duplicate the fields throughout for that medical school. I next describe the sources of the data on medical schools and the process used to merge and construct the fields.

NIH Funding Data

The National Institutes of Health organizes the data on its expenditures and makes it available through RePORT. The records of each project funded by the NIH is available for download through http://projectreporter.nih.gov/reporter.cfm. The records identify the projects by an application id and fields include the institution type, total cost and project categories. I include funding for projects designated to Schools of Medicine, Schools of Medicine and Dentistry, and Overall Medical as these categories were the major categories at which the recipient was affiliated with an MD medical school. I wished to include funding only for extramural and cooperative research activities, and training and fellowship programs funded by the NIH in a medical school. So, I dropped activity codes beginning with G, C, H as these were designated for construction, resource development and community service. Further, I dropped activity codes beginning with N and Z since those data are available only after 2007.

I used the records from all project costs incurred in the financial years 2000 to 2010 that satisfy the criteria above and aggregated the project costs to the organization name. I wish to construct the average annual NIH research costs incurred at these medical schools during this period. I infer that a school was operating during a given year if it secured some NIH funding. All but thirteen schools secured NIH funding during each of the eleven years in the sample. Six schools did not receive any NIH funds during this period even though they were operating (as indicated by online sources) and their eleven year annual average NIH costs were set to zero. For the remaining seven medical schools, I established the number of years the school was operating by searching for the history of the school from the history of the medical college published on their websites.

These data were merged with the data from the National GME Census using the medical school names. Of the 135 MD medical schools in the GME Census, 129 medical schools were matched successfully to a counterpart in the NIH funding data. I verified that the remaining six schools did not have any records in project RePORT in the categories considered.

Medical School Admission Requirements (MSAR)

I used the records from the 2010-2011 MSAR publication of the AAMC to augment the medical school characteristics with the state and the median MCAT score of the admits into a medical school. The merge was done using the medical school name and MCAT score data was found for all but seven of the 135 MD granting medical schools. Data on the state the medical school is located in was found for all MD medical schools.

B.5.3 Medicare Data

Here, I describe the merge and construction of the Case Mix Index and Wage Index variables. The instrument, based on Medicare reimbursement rates is described in Section B.7.

I use the records from the Medicare provider files to construct the variables primary care reimbursement rates, the Medicare wage index and the case mix index. The institution ids for all affiliates were merged with Medicare provider identifiers by the name of the provider by using the 1997 PPS files, and then using the 2010 Impact Files. A second check was conducted for primary institutions of the programs, and for affiliates when primary institutions were not matched to Medicare data. In a small number of instances, there are multiple matched CMS identifiers for a single institution. Medicare variables were averaged across these multiple matches.

Medicare Wage Index and Case Mix Index

The Center for Medicare Services calculated a Wage Index and Case Mix Index for each provider.[8] I merged the CMS data with primary institution. In a small number of instance, the primary institution did not have a match with Medicare data. In these cases, I calculated the average of the variable for all affiliates with Medicare data. In a total of 63 out of 3441 cases, the case mix index was not available even for affiliates. Here, in the structural estimates, I used an imputed value from a linear regression on all other characteristics included in the demand system. Finally, missing values of the wage index were imputed using the geographic definitions Medicare uses to calculate the wage index.

B.5.4 Identifying Rural Programs

I use two sources of data to identify the set of rural family medicine program.

1. The American Academy of Family Physicians has a program directory of all family medicine programs in the United States. The program directory lists the community setting of the program as one or more of Urban, Suburban, Rural, Inner-city. Programs for which only rural was listed as the community setting are considered rural programs by this definition. The records from this directory were scraped on 01/05/2012. I manually merged the set of rural programs to AMA data using the name of the program, the hospital and the street address. In the years 2003-2004 to 2010-2011, this procedure identified 438 program-year observations as rural programs.

[8] The files and the description of the calculation for the wage index is given on http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/wageindex.html and the Case Mix Index is described on http://en.wikipedia.org/wiki/Case_mix_index

2. The program names in the AMA data often directly indicate whether a program is a rural program or not. For instance, the University of Wisconsin sponsors several programs in family medicine, one of which is named ”University of Wisconsin (Madison) Program” and the other named ”University of Wisconsin (Baraboo) Rural Program.” I consider all programs with rural in the name during the same period of the program as a rural program. This procedure identified a total of 159 program-year observations as rural programs in the years 2003-2004 to 2010-2011, of which a total of 115 program-year observations overlapped with program-year observations identified as a rural program using the previous procedure.

In 2010-2011, I checked for contradictions where a program with rural in the program name listed a community setting other than rural in the AAFP directory. There were a total of 5 programs that were classified as rural according to rule 2 but not rule 1. Of these, in four cases, the program directory did not have any information other than the name and address of the program. The community setting for the remaining program was listed as suburban as well as rural.

B.5.5 Resident Birth Location

The birth location of the resident is recorded as city, state and country code. The following steps were carried out to improve the quality of the data and then to identify whether a resident was born in a rural location in the United States:

1. I convert the AMA country identifiers, which are not unique across years, to the corresponding ISO 3166-1 alpha-3 identifier using the country name provided by the AMA. Except for some former soviet nations and territories of the UK, US and Netherlands, a unique match was available.

2. The state and country for observations with only the city name were imputed using the state and country for an identically spelled city if that state-country combination constituted more than 50% of the observations for that city. This imputation was carried out using the GME Census data from 1996-1997 to 2010-2011 in five specialties: internal medicine, pediatrics, OB/GYN, pathology and family medicine.

3. For US born residents, city-state combinations were geocoded. The observations for which the geocoder indicated a match with unexpected accuracy (more than, or less than city level accuracy) were checked by hand and minor spelling errors were corrected. The corrections were put through the geocoder for a second time. Ambiguous entries were coded as missing data.

4. The county of birth for US born residents was extracted matched with a list of counties that belong to a Metropolitan Statistical Area in order to construct the rural birth indicator.

B.5.6 Other Data

CPI-U

I downloaded the records of the monthly Consumer Price Index for All Urban Consumers from the Federal Reserve Economic Data (FRED) website. I use the December observation for the CPI-U for a year.

Rent

Census Data from the 2000 US Census was downloaded from nhgis.org. I used county level aggregates from sample file 1 for population, age and race variables, and from sample file 2 for income and rent variables. The median gross rent is used as the measure of rent as it adjusts for the utility payments.

The 2010 US Census did not use the long form on which data on the rent paid is collected. Consequently, data on the county level median gross rent was downloaded from the 2006-2010 American Community Survey using Social Explorer. These rent numbers are adjusted to 2010 dollars by Social Explorer. The five-year aggregate was preferred to the annual or three-year aggregates since the latter did not cover all counties in the US.

To construct the median gross rent variable, I convert the median rent data from the 2000 US Census into 2010 dollars by using CPI-U. A linear interpolation between the 2000 and 2010 rent data for the interim years.

Merging The city, state and zip code of the program and institutions were used to geocode the latitude and longitude of the zip code's centroid. These latitudes and longitudes were then used to determine the county in which the program or institution is located using county shape files provided by NHGIS. The geographic ids from this process were used to merge these with the data files. Every program in the sample was successfully matched in this process.

B.5.7 Miscellaneous Issues

1. For the preference estimates, imputation of salaries for missing data was done for 23 observations out of 3441 using a linear regression on the other characteristics included in the model.

2. The program survey asks for the number of first year positions offered in the next academic year. I use this as the preferred measure of the program's capacity when available. In ten instances, this field was not available and for nine of these instances, it was imputed from the value of the field from the previous year. In the remaining instance, the number of first year residents in the program was taken to be the number of positions offered. I checked to ensure that the reported number positions offered next year is equal to the number of matched than the value of the field from the previous year.

   I find instances when the number of residents in first year positions exceeds this capacity measure. In these cases, I take the maximum of the number matched to the program and the lagged response to the first year enrollment as the program's capacity. In more than 75% of the cases, the number matched did not exceed the reported number of positions by more than one. Table B.7 summarizes the number of observations affected by this change and the mean size of the change. One reason for the discrepancy may be residents that repeated their first year training or deferred enrollment.

Table B.7: Capacity Adjustments

Year	Number of program capacities adjusted	Average adjustment	Maximum adjusment
2003-2004	51	1.25	3
2004-2005	53	1.32	5
2005-2006	72	1.32	4
2006-2007	57	1.14	2
2007-2008	74	1.35	5
2008-2009	67	1.40	4
2009-2010	65	1.35	5
2010-2011	71	1.54	6

Notes: Capacities are adjusted upwards only. Average adjustment is reported conditional on adjustment.

B.6 The Distribution of Physician Starting Salaries

The experience adjustment uses the following Mincerian wage regression to capture the impact on physician productivity:[9]

$$\ln y_i = \rho_0 + \rho_1 t_i + \rho_2 t_i^2 + \rho_c c_i + e_i. \tag{B.9}$$

Here, y_i is the earnings of physician i, t_i is the experience of physician i, c_i is a vector of controls and e_i is mean zero error. The functional form is motivated by a multiplicative return to human capital, which increases with job experience up to a maximum before depreciating.

I use records from the restricted-use file of family practice physicians from the Health Physician Tracking Survey of 2008 to estimate ρ. The survey collects data on the income category of physicians in the United States, with medical specialty, years practicing medicine and a variety of other fields related to their medical practice. The survey asks for the income earned by the physician in 2006 from medically related activities, excluding returns on investments in stocks or assets in their practice. The income field is coded into groups \$50,000, with the lowest category for physicians with an income under \$100,000

[9]See motivating theoretical model in Ben-Porath (1967), some early empirical work in Mincer (1974). Thomas Lemieux (2006) and Heckman et al. (2003) survey the literature on mincer regressions.

and the highest category for physicians with an income of $300,000 or more. I use an interval regression in which $e_i \sim N\left(0, \sigma_e\right)$ to estimate $\left(\rho,\ \sigma_e\right)$.

Table B.8 presents summaries from the subpopulation of physicians under the age of 60 in 2006, the year of the income data in the survey. The vast majority of family physicians are salaried and earn $200,000 or less. Table B.9 presents maximum likelihood estimates from the interval regression model. The point estimates evidence for concavity in returns to experience and a gender-pay gap that is well-documented in the empirical literature. A comparison of estimates in columns (2) and (3) also suggest some heteroskedasticity in the distribution of pay across experience levels. Column (4) estimates a quadratic functional form for this heteroskedasticity and finds a concave relationship, with a higher cross-sectional variation in earnings for physicians in the middle of their career than for physicians early or late in their career.

Table B.8: Characteristics of Family Medicine Doctors in the US

	Mean	Std. Dev
Observations	698	
Income less than \$100K	16.64%	
Income between \$100K to \$150K	35.43%	
Income between \$150K to \$200K	27.76%	
Income between \$200K to \$250K	9.95%	
Income between \$250K to \$300K	6.36%	
Income more than \$300K	3.86%	
Income Type: Hourly	4.48%	
Income Type: Salary	71.73%	
Income Type: Profits from Practice	23.79%	
Hours Last Week	50.19	13.75
Weeks Worked	47.48	4.63
Full Time	87.95%	
Experience	13.69	8.35
Foreign Medical Graduate	15.17%	
Female	30.83%	
Practice Type: Solo/Two Physician	31.82%	
Practice Type: Group	46.27%	
Practive Type: Other	21.91%	
Large Metropolitan Area	46.89%	
Small Metropolitan Area	32.44%	
Non-Metropolitan Area	20.67%	

Notes: Sample of Family Practice Physicians in the Health Tracking Physician Survey of 2006 with non-missing income, starting medical practice in or before 2006. Income from medically related activities in 2006. Hours reported for medically-related activities. Income excludes in returns from investments in financial and medical capital. Experience defined as number of years since beginning medical practice. Full-time defined as more than 35 hours spent on medical activities and more than 40 weeks worked in 2006. Large Metropolitan Area has more than 1 million residents.

Table B.9: Income of Family Medicine Doctors

Dependent Variable	Log Income from Practice					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Interval Regression Estimates						
Experience	0.0144*	0.0124	0.0819**	0.0117	0.0147*	0.0126
	(0.0070)	(0.0070)	(0.0256)	(0.0073)	(0.0069)	(0.0069)
Experience-squared	-0.0003	-0.0004	-0.0063**	-0.0004	-0.0005	-0.0004
	(0.0002)	(0.0002)	(0.0024)	(0.0002)	(0.0002)	(0.0002)
Female		-0.2617***	-0.2621***	-0.2121***	-0.2581***	-0.2759***
		(0.0352)	(0.0421)	(0.0372)	(0.0346)	(0.0347)
Foreign Medical Graduate						-0.0446
						(0.0441)
Practice Type: Solo/Two Physician						-0.1392***
						(0.0365)
Practice Type: Other						0.0062
						(0.0345)
Small Metropolitan Area						0.0544
						(0.0347)
Non-Metropolitan Area						0.0647
						(0.0398)
Contant	11.7822***	11.9014***	11.7658***	11.9388***	11.8955***	11.9228***
	(0.0413)	(0.0449)	(0.0596)	(0.0465)	(0.0433)	(0.0500)
Heteroskedascitiy by experience					Y	
Sample			$<$ 10 years exp	Full-time		
Oberservations	698	698	295	616	698	698
Total Sample Weight	60620	60620	25612	53318	60620	60620
Panel B: Estimated Distribution Statistics at Zero Experience						
Mean	153660.74	148524.83	127612.97	157920.84	144746.55	146895.78
Std. Dev.	68769.35	63368.95	47622.36	65432.74	50911.87	61416.25

Notes: Interval regressions. Statistics at zero experience, means of other characteristics.

B.7 Medicare Reimbursement Rates and Instrument Details

B.7.1 Description of Medicare Reimbursement Regulations

Medicare Direct Graduate Medical Expenditure (DGME) payments are designed to compensate teaching hospitals for expenses directly incurred due to the training of residents. The methodology used to determine these payments was established in the Consolidated Omnibus Budget Reconciliation Act (COBRA) of 1985, and are implemented as per 42 CFR §§ 413.75 to 413.83. Here, I provide a broad outline of the method used to determine Medicare DGME payments and the PCPRA variable used in the analysis.

Roughly, the total DGME reimbursements to a hospital is the product of the hospital specific per resident amount (PRA), the weighted number of full-time equivalent residents (FTE) and Medicare's share of total inpatient days. The PRA is determined using the total costs of salaries and fringe benefits of residents, faculty and administrative staff of the residency program and allocated institutional overhead costs divided by the total number of full time equivalent residents in a base year, usually 1984 or 1985. Hospitals that began sponsoring residency training after 1985 were grandfathered into the program using their first year of reported costs as the base year. After 1997, a new hospital's per resident amount was based on the reported costs of other programs in the geographic area, which is an MSA/NECMA, rest of state or a census division depending on the number of other providers sponsoring GME. The Balanced Budget Act of 1997 also introduced certain ceilings and floors on the per resident amount. See Gentile Jr. and Buckley (2009) for a more comprehensive legislative history of Medicare reimbursement of Graduate Medical Education.

Between 1985 and 2000, the PRA for a hospital was revised by adjusting for the 12 month change in CPI-U, and minor changes on previously misallocated costs. An exception was made in 1993 and 1994 when two separate PRAs were effectively created, one for primary care and obstetrics and gynecology residents and the other for all other residents. In these two years, the non-primary care PRA was not adjusted for inflation.

Subsequent to 2000, the per resident amounts were also adjusted using the change in

CPI-U but were subject to a floor and ceiling put in place by the The Balanced Budget Act of 1997. The floor increased the PRAs of hospitals that were below 70% of the (locally-adjusted) national average per-resident amount to 70% of the total and later to 85%. The ceiling gradually decreased the PRAs of hospitals that were above 140% of the (locally-adjusted) national average per-resident amount until the PRA of a hospital fell below the ceiling. The exact procedure used to make these adjustments is detailed in 42 CFR § 413.77. The Balanced Budget Act of 1997 also created new regulations on the manner in which the number of full-time equivalent residents was determined. These regulations are detailed in 42 CFR § 413.86.

B.7.2 The Instrument: Competitor Reimbursement Rates

To construct competitor reimbursements, I first extract the records from the fields ”Updated per resident amount for OB/GYN and primary care” and ”Number of FTE residents for OB/GYN and primary care” on lines 2 and 1 respectively in form CMS-2552-96, Worksheet E-3, Part-IV for the cost reporting period beginning October 1, 1996 and before September 30,1997. As per the instructions for this form (3633.4), this is the latest period for which the response to the field was required by the hospitals. Indeed, I found only five observations for this field in the cost reporting period ending October 1, 1998 and no observations in the next period. The per resident amount variable is recorded in cents, and so is first converted into dollars. Both fields were winsorized at the bottom at top 1 percent since the range of values were extreme. Barring the effects of winsorizing the data, the distribution of the per resident amount variable is similar to Figure B.1 taken from Newhouse and Wilensky (2001). While some institutions have per resident amounts less than $40,000, others are reimbursed at rates higher than $200,000.

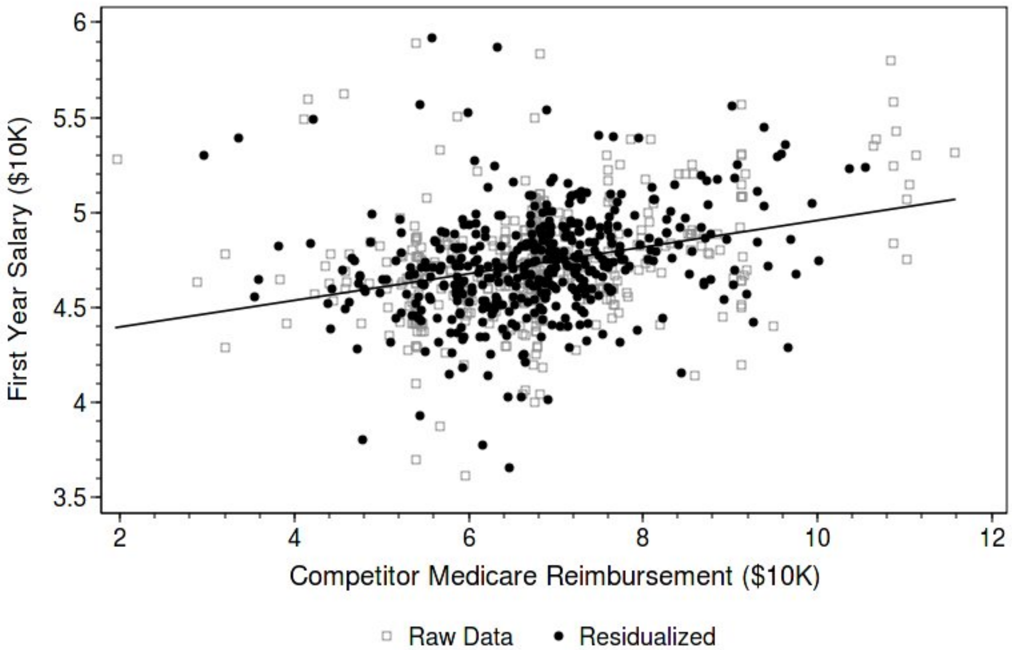
The Competitor Reimbursement variable for an institution is constructed in order to mimic the per resident amount calculation done by Medicare for new sponsors. As given in equation (2.7), the (weighted) Competitor Reimbursement variable for a program is the average (weighted by FTE) of all primary care per resident amounts in the primary institution's geographic area (MSA/NECMA or the rest of the state) other than that of the primary institution. When this average is constructed from less than three observations, the census division is used. This variable is then merged to the primary institution of a program as defined earlier.

Figure B.3 depicts the state-averaged variation in the instrument that is not explained by the controls included in the preference estimates and a program's own reimbursement rate. A degree of spatial correlation within a census division is noticeable due to the definition of the geographical units used. Table B.10 presents regressions of the instrumental variable on characteristics included in the preference estimation, as well as location characteristics such as median age, median household income, crime rates, total population and college share. These location characteristics, together with program characteristics explain only 27% of the variation in the instrument. Strictly speaking a test for exogeneity with respect to the additional location characteristics would be rejected at the 1% level. However, the location characteristics together explain only about 6% of the variation not explained by the other controls that are included in the preferences estimates. Columns (4-6) show that characteristics of the program itself explain about 35% of the variation in its reimbursement rates and the addition of location characteristics is not important. These findings are consistent with Anderson (1996), which argues against this reimbursement schemes on the basis that other cost predictors do not correlate very strongly with per resident amounts. Strictly speaking, these findings do not fully support strict exogeneity of the instrument.

Table B.10: Medicare Reimbursement Rates on Characteristics

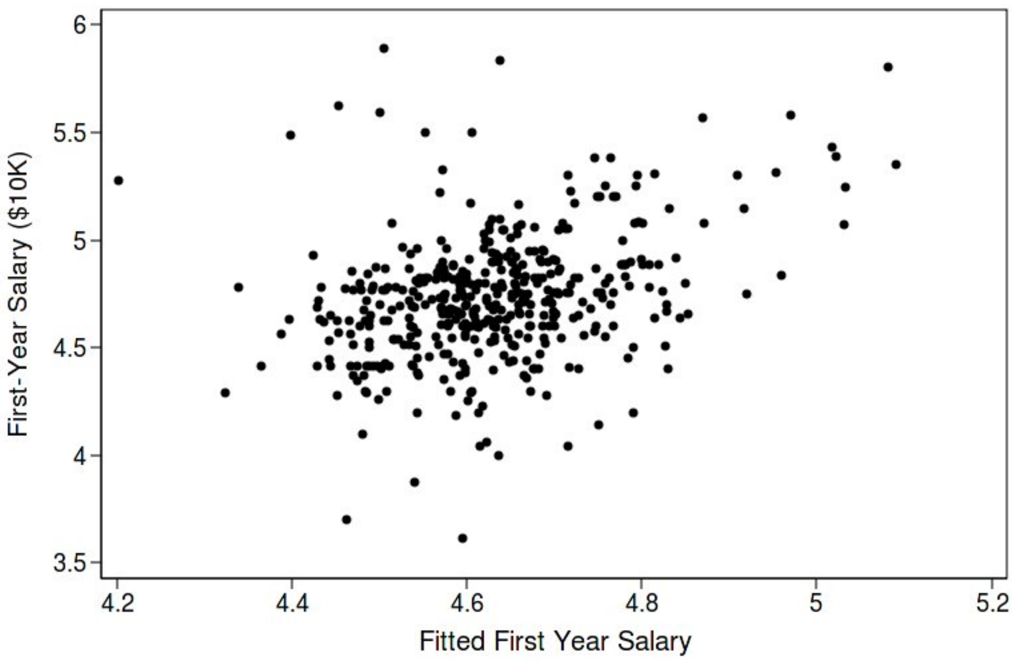
Dependent Variable	Log Competitor Reimbursements			Log Reimbursements		
	(1)	(2)	(3)	(4)	(5)	(6)
Log Rent	-0.0057	-0.0282	-0.1746**	-0.0004	-0.2023	-0.1330
	(0.0632)	(0.0737)	(0.0879)	(0.1219)	(0.1579)	(0.1973)
Log Wage Index	0.3924***	0.3425***	0.0937	0.7497***	0.6509***	0.4406
	(0.1036)	(0.1038)	(0.1042)	(0.1977)	(0.2042)	(0.2679)
Log Reimbursement	0.1701***	0.1538***	0.1410***			
	(0.0227)	(0.0221)	(0.0232)			
Location Characteristics		Y	Y		Y	Y
Small Cities (< 3 mi in Population)			Y			Y
Observations	3,441	3,441	2,407	3,441	3,441	2,407
R-squared	0.2335	0.2934	0.2719	0.3528	0.3731	0.3550

Notes: Linear regressions. Location characteristics include median age (county), log median household income (county), log total population (MSA/county), violent crime and property crime rates from FBI's Crime Statistics/UCR (25 mi radius weighted by 1/distance), dummies for no data in that radius and log college share (MSA/rest of state). All columns include a constant term, log # beds, log NIH Fund (Major), log NIH Fund (Minor), Log Case Mix Index, Program Type Dummies, Rural Program Dummy and dummies for programs with no NIH funding at major affiliates, for no NIH funding at minor affiliates, and a dummy for missing Medicare ID at program institutions. The Competitor Reimbursement is a weighted average of the Medicare primary care per resident amounts of institutions in the geographic area of a program other than the primary institutional affiliate of the program. Geographic area defined as in Medicare DGME payments: MSA/NECMA or Rest of State unless less than 3 other observations constitute the area, in which case the census division is used. See data appendix for description of variables and details on the construction of the reimbursement variables. For columns (4-6), a program's reimbursement rate is truncated below at $5,000 and a dummy for these 46 truncated observations is estimated as well. Standard errors clustered at the program level in parenthesis. Significance at 90% (*), 95% (**) and 99% (***) confidence.



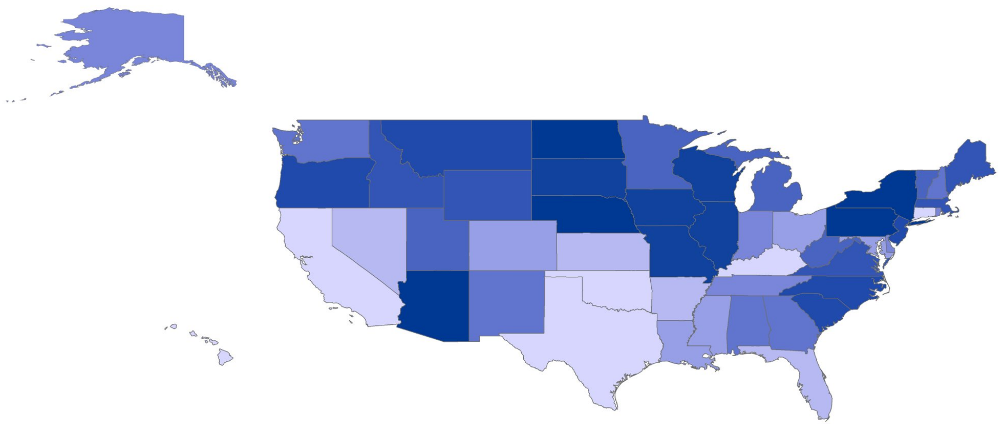
Notes: Sample restricted academic year 2010-2011. To construct the residualized scatter plot, I first regressed the X-axis and Y-axis variables on County Median Rent (Gross), Rural Program, Medicare Wage Index, Log NIH Fund (Major), Log NIH Fund (Minor), Log # Beds, Medicare Case-Mix Index and dummies for No NIH Fund (Major), No NIH Fund (Minor), missing Medicare ID. The X-axis and Y-axis residuals estimated from these regressions are scattered.

Figure B.1: Relationship Between Wages and Competitor Reimbursements



Notes: To construct the fitted salaries, I regressed the First Year Salary on Competitor Reimbursements, County Median Rent (Gross), Rural Program, Medicare Wage Index, Log NIH Fund (Major), Log NIH Fund (Minor), Log # Beds, Medicare Case-Mix Index and dummies for No NIH Fund (Major), No NIH Fund (Minor), missing Medicare ID. The regression was estimated on the full sample from the academic years 2002-2003 to 2010-2011. The scatter plot shows the salaries and fitted values from the academic year 2010-2011 alone. The Competitor Reimbursement is a weighted average of the Medicare primary care per resident amounts of institutions in the geographic area of a program other than the primary institutional affiliate of the program. Geographic area defined as in Medicare DGME payments: MSA/NECMA unless less than 3 other observations constitute the area, in which case the census division is used. See data appendix for description of variables and details on the construction of the reimbursement variables.

Figure B.2: Heteroskedasticity in First Stage Residuals



Notes: Average residuals of the Competitor Medicare Reimbursements by state. Colors categorized by 10 equally sized quantiles with darker colors indicating higher values. Program sample restricted academic year 2010-2011. To construct the average residuals by state, I first regressed Competitor Medicare Reimbursements on County Median Rent (Gross), Rural Program, Medicare Wage Index, Log NIH Fund (Major), Log NIH Fund (Minor), Log # Beds, Medicare Case-Mix Index and dummies for No NIH Fund (Major), No NIH Fund (Minor), missing Medicare ID. The estimated from these regressions were averaged by the state a program is located in. The Competitor Reimbursement is a weighted average of the Medicare primary care per resident amounts of institutions in the geographic area of a program other than the primary institutional affiliate of the program. Geographic area defined as in Medicare DGME payments: MSA/NECMA unless less than 3 other observations constitute the area, in which case the census division is used. See data appendix for description of variables and details on the construction of the reimbursement variables.

Figure B.3: Geographic Distribution of Competitor Reimbursements

Appendix C

Appendices to Chapter 3

C.1 Proofs: Identification

C.1.1 Proof of Lemma 3.1

As noted in Remark 3.1, the q-th quantile of each side matches with the q-th quantile of the other. If F_U and F_V are the distributions of utilities on each side, the match operator on utilities, which determines the firm quality index u that a worker of quality v is matched with, is given by $F_U^{-1} \circ F_V$. Note that these are both monotonically increasing.

Define T_ε be a convolution with F_ε :

$$T_\varepsilon \left(F_{h(x)}\right)(u) = \int_{-\infty}^{\infty} F_{h(x)}\left(u - \varepsilon\right) dF_\varepsilon \tag{C.1}$$

and T_η be a convolution with F_η and let $T_\cdot^{-1}$ be the associated inverse transforms. Note that T_ε is strictly increasing (wrt the partial order induced by First Order Stochastic Dominance) in $F_{h(x)}$. Further, $T_\cdot^{-1}$ exists since η admits a density and is strictly increasing since it is the inverse of a monotone operator.

The distribution of U that $h(x)$ is matched with is then given by $F_U^{-1} \circ F_V \circ T_\varepsilon \left(\delta_{h(x)}\right)$ where $\delta_{h(x)}$ is the dirac delta function at $h(x)$. This implies that the distribution on $g(z)$ that a given $h(x)$ is matched with is $T_\eta^{-1} \circ F_U^{-1} \circ F_V \circ T_\varepsilon \left(\delta_{h(x)}\right)$. Since a composition of strictly increasing functions is strictly increasing, $T_\eta^{-1} \circ F_U^{-1} \circ F_V \circ T_\varepsilon \left(\delta_{h(x)}\right)$ is strictly increasing in $\delta_{h(x)}$.

Now consider two values $x_1,\, x_2 \in \operatorname{supp} m^x$ such that $\mu(x_1,\cdot) = \mu(x_2,\cdot)$ where μ is the stable match. Assume, wlog that $h(x_1) > h(x_2)$. This implies that $T_\eta^{-1} \circ F_U^{-1} \circ F_V \circ T_\varepsilon\left(\delta_{h(x_1)}\right) >_{FOSD} T_\eta^{-1} \circ F_U^{-1} \circ F_V \circ T_\varepsilon\left(\delta_{h(x_2)}\right)$. Therefore, it must be that $\mu(x_1,\cdot) \neq \mu(x_2,\cdot)$ since the push forward $\mu_{\#g}(x_1,\cdot) \neq \mu_{\#g}(x_2,\cdot)$.

Conversely, if $h(x_1) = h(x_2)$, we have that the distributions of $g(z)$ matched with x_1 and x_2 are the same: $T_\eta^{-1} \circ F_U^{-1} \circ F_V \circ T_\varepsilon\left(\delta_{h(x_1)}\right) = T_\eta^{-1} \circ F_U^{-1} \circ F_V \circ T_\varepsilon\left(\delta_{h(x_2)}\right)$. Assumption 3.3 implies that the distribution of z that x_1 and x_2 are matched with are also identical, i.e. $\mu(x_1,\cdot) = \mu(x_2,\cdot)$.

Hence, $h(x_1) = h(x_2)$ if and only if $\mu(x_1,\cdot) = \mu(x_2,\cdot)$, which is known from the data. A symmetric argument identifies the level sets of $g(\cdot)$.

C.1.2 Proof of Proposition 3.3

Identification of h up to a positive monotone transformation follows immediately from Proposition 3.1 and the assumption that $\frac{\partial h(\cdot)}{\partial x_k} > 0$.

In the proof of Proposition 3.1, although the argument was stated for h, we can replicate it to conclude that $g(z_1) > g(z_2)$ implies that $T_\varepsilon^{-1} \circ F_V^{-1} \circ F_U \circ T_\eta\left(\delta_{g(z_1)}\right) >_{FOSD} T_\varepsilon^{-1} \circ F_V^{-1} \circ F_U \circ T_\eta\left(\delta_{g(z_2)}\right)$ i.e. the distribution of $h(x)$ matching with z_1 dominates the distribution for z_2. This notion of dominance is invariant to positive transformations of $h(x)$. Hence, we can order the level sets of $g(\cdot)$.

C.1.3 Proof of Proposition 3.4

Assume that the function $h(\cdot)$ is known up to a positive monotone transformation and let $\tilde{x}_j = \tilde{h}(x_j)$ for a particular, twice continuously differentiable function with identical level sets as $h(\cdot)$. Let $F_{\tilde{X},Z}$ be the joint distribution of $\tilde{X}$ and Z observed in the data generated by the model defined by equations (3.1) and (3.2). Assumption 3.3 implies that this a sufficient statistic for $F_{\tilde{X},Z}$.

The proof proceeds by rewriting the matching model with $\varepsilon \equiv 0$ in terms of the transformation model of Chiappori and Komunjer (2008). We appeal to Chiappori and Komunjer (2008), Proposition 1 stating that the transformation model is correctly specified. We need to re-write this matching model and verify Assumptions A1-A4 in Chiappori and Komunjer (2008).

Remark 3.1 implies that

$$\begin{aligned} F_{\tilde{X}|Z} &= \int h^{-1}\left(F_U^{-1}F_V\left(g\left(Z\right)+\eta\right)\right)dF_\eta \\ &= \int \Gamma\left(g\left(Z\right)+\eta\right)dF_\eta. \end{aligned}$$

Since F_U^{-1}, F_V, h^{-1} are twice continuously differentiable, so is $\Gamma = h^{-1}\left(F_U^{-1}F_V\left(\cdot\right)\right)$. This verifies Assumption A1. Assumption A2 follows since Γ and $g\left(\cdot\right)$ may be scaled in order to ensure that $\int f_\varepsilon\left(t\right)dt = 1$. Assumption A3 is stronger than the Assumption 3.1 (ii) in this paper. Assumption A4 in follows from twice continuous differentiability of g.

C.1.4 Lemma C.2: Preliminary for Propositions 3.6 and 3.5

Let $v = h\left(x\right)+\varepsilon$, where $h\left(x\right)$ is strictly increasing with $h\left(\bar{x}\right) = 0$, $h'\left(\bar{x}\right) = 1$ and let ε be median zero with density f_ε. For quantile $\tau \in [0,1]$, let $f_{\tau|x}\left(\tau\right)$ be the density on $v = F_V^{-1}\left(\tau\right)$ given x, where $F_V\left(v\right) = \int F_\varepsilon\left(v-h\left(x\right)\right)dF_X$. i.e. $f_{t|x}\left(x\right) = f_\varepsilon\left(F_V^{-1}\left(\tau\right)-h\left(x\right)\right)$.

Lemma C.2. *The function $h\left(x\right)$ and the density f_ε are identified from $f_{\tau|x}\left(\tau\right)$ and F_X if $h\left(x\right)$ is differentiable.*

Proof. Let $\phi\left(x,x'\right)$ be the probability that $h\left(x\right)+\varepsilon > h\left(x'\right)+\varepsilon'$ given x and x'. $\phi\left(x,x'\right)$ is identified from $f_{\tau|x}\left(\tau\right)$ since it can be written as

$$\phi\left(x,x'\right) = \int_0^1 \int_{\tau>\tau'} f_{\tau|x}\left(\tau\right) f_{\tau|x}\left(\tau'\right) d\tau d\tau'.$$

However, $\phi\left(x,x'\right)$ can also be written in terms of F_X and f_ε as

$$\phi\left(x,x'\right) = \int F_\varepsilon\left(h\left(x\right)+\varepsilon-h\left(x'\right)\right) f_\varepsilon\left(\varepsilon\right) d\varepsilon.$$

Taking the derivative with respect to x and x', we get

$$\begin{aligned} \frac{\partial\phi\left(x,x'\right)}{\partial x} &= h'\left(x\right)\int f_\varepsilon\left(h\left(x\right)+\varepsilon-h\left(x'\right)\right) f_\varepsilon\left(\varepsilon\right) d\varepsilon \\ \frac{\partial\phi\left(x,x'\right)}{\partial x} &= -h'\left(x'\right)\int f_\varepsilon\left(h\left(x\right)+\varepsilon-h\left(x'\right)\right) f_\varepsilon\left(\varepsilon\right) d\varepsilon. \end{aligned}$$

The ratio $\frac{\partial\phi(x,x')}{\partial x} / \frac{\partial\phi(x,x')}{\partial x}$ is identified and is equal to

$$-\frac{h'(x)}{h'(x')}.$$

Since $h'(\bar{x})$ is known, $h'(x)$ can be determined everywhere. The boundary conditions $h(0) = 1$ provides the unique solution to the resulting differential equation determining $h(\cdot)$.

We now need to show that F_ε is identified. Let $R_x(t)$ be the (utility-) rank distribution of x, i.e. the probability that the utility of x is below the t-quantile in utility distributions, $\mathbb{P}\left(h(x) + \varepsilon \le F_V^{-1}(t)\,|x\right)$. $R_x(\tau)$ is known since it is equal to $\int_0^t f_{\tau|x}(\tau)\,d\tau$. Since F_V^{-1} is continuous and ε admits a density, $R_x(t)$ is continuous and strictly increasing in t. Let τ^* be the median rank of $\bar{x}$, i.e. $R_{\bar{x}}(t^*) = \frac{1}{2}$. Since ε is median-zero and $h(\bar{x}) = 0$, the median $\bar{x}$ has $v = 0$. For any x, $R_x(t^*)$ is therefore the probability that $h(x) + \varepsilon \le 0$ given x, i.e. $R_x(t^*) = F_\varepsilon(-h(x))$. Since $h(x)$ has full support on $\mathbb{R}$, F_ε is identified. $\square$

C.1.5 Proof of Proposition 3.5

Since $g(z)$ is identified upto a positive monotone transformation (Proposition 3.3), we can treat z as a scalar without loss of generality.

Under Assumption 3.3, for a firm with traits (z_j, η_j), the distribution of workers $F^W_{h(X)|(z,\eta)}$ at depends only on u, $F^W_{h(X)|u=g(z_j)+\eta_j}$, and is given by $T_\varepsilon^{-1} \circ F_V^{-1} \circ F_U(u_j)$ where $u_j = g(z_j) + \eta_j$, and T_ε, F_V and F_U are as defined in the proof of Lemma 3.1. Since T_ε, F_V and F_U are monotone, $T_\varepsilon^{-1} \circ F_V^{-1} \circ F_U(u_j)$ is increasing with respect to first order stochastic dominance. Hence, $F^W_{h(X)|u_j=g\left(z_{j'}\right)+\eta_{j'}} <_{FOSD} F^W_{h(X)|u_j=g(z_j)+\eta_j}$ if and only if $g(z_{j'}) + \eta_{j'} < g(z_j) + \eta_j$. The distributions, $F^W_{h(X)|u}$ are therefore totally ordered. For any $\tau \in [0,1]$, let $F^W_{h(X)|\tau} = F^W_{h(X)|F_U^{-1}(\tau)}$ be the τ-th quantile of the family of distributions $\left\{F^W_{h(X)|u}\right\}_{u\in R, u\sim F_U}$. The data consists of a measure over the worker distributions at firms F^W_X, and since $h(\cdot)$ is known, we may identify the measure over distributions of observable worker quality at firms, $F^W_{h(X)}$. Hence, for every $\tau \in [0,1]$, the quantity $F^W_{h(X)|\tau}$ is identified: it the worker distribution that stochastically dominates the worker distribution at precisely a fraction τ of firms. $F^W_{h(X)|\tau}$ can be used to determine

$F_{\tau|Z}(\tau) = \mathbb{P}\left(g(Z) + \eta < F_U^{-1}(\tau)\,|Z = z\right)$ for any τ, since

$$\mathbb{P}\left(g(Z) + \eta < F_U^{-1}(\tau)\,|Z = z\right) = \mathbb{P}\left(F^W_{h(X)|g(Z)+\eta} <_{FOSD} F^W_{h(X)|\tau}|Z = z\right).$$

Finally, note that $F_{\tau|Z}(\tau)$ is differentiable in τ since F_U admits a density. Hence, $f_{\tau Z}(\tau)$ is known. Lemma C.2 implies that $g(z)$ and f_η are identified.

C.1.6 Proof of Proposition 3.6

Let $f_{X|t}$ be the conditional density of worker traits X that a firm of observable quality $g(z) = t$ is matched with. This density is identfied from F_{XZ} and the function $g(\cdot)$. We first show that, $\tilde{f}_{X|u}(x)$, the density of firm traits that firms of true desirability u are matched with is identified.

Under Assumption 3.3, $f_{X|q}$ is given by

$$f_{X|q}(x) = \int_{-\infty}^{\infty} \tilde{f}_{X|u=t+\eta}(x)\, f_\eta(\eta)\, d\eta.$$

Denote $f_{X|q}(x)$ with $\psi_x(t)$ and $\tilde{f}_{X|u}(x)$ with $\tilde{\psi}_x(u)$, and re-write

$$\psi_x(t) = \int_{-\infty}^{\infty} \tilde{\psi}_x(t+\eta)\, f_\eta(\eta)\, d\eta.$$

Substituting $u = t + \eta$, we can rewrite $\psi_x(t)$ as the following Fredholm equation of the first kind

$$\psi_x(t) = \int_{-\infty}^{\infty} \tilde{\psi}_x(u)\, f_{\eta^*}(t-u)\, du$$

where $f_{\eta^*}(t-u) = f_\eta(u-t)$. Let $\mathcal{F}_t[\psi_x(t)]$ be the fourier transform of $\psi_x(t)$, we can solve for $\tilde{\psi}_x(u)$ as

$$\tilde{\psi}_x(u) = \int_{-\infty}^{\infty} \frac{\mathcal{F}_t\left[\tilde{\psi}_x(t)\right](\omega)}{\mathcal{F}_t\left[f_{\eta^*}(t)\right](\omega)} \exp^{2\pi i\omega t}\, d\omega$$

since $\mathcal{F}_t[f_{\eta^*}(t)](\omega)$ is known and non-zero everywhere since f_η has a known, non-vanishing characteristic.

Since F_U, F_X and $\tilde{f}_{X|u}$ are known, $\tilde{f}_{U|x}$ the density of firm quality u that workers with observable type x are matched with can be determined from Bayes' rule :

$$\tilde{f}_{U|x}(u|X = x) = \frac{\tilde{f}_{X|u}(x|U = x)\, f_U(u)}{f_X(x)}.$$

Therefore, for any quantile $\tau \in [0,1]$, the density that a worker with observed characteristic x is matched with a firm in the τ-th quantile of the firm quality distribution, $\tilde{f}_{\tau|x}\left(F_U^{-1}(\tau)\,|X=x\right)$, is known. Quantile-quantile matching implies that $\tilde{f}_{\tau|x}\left(F_U^{-1}(\tau)\,|X=x\right) = f_{\tau|x}(\tau)$, the density that a worker with characteric x has τ-th quantile of the worker quality distribution. By Lemma C.2, $h(x)$ and $f_{\varepsilon}(\varepsilon)$ are identified.

C.2 Proofs: Estimation

C.2.1 Preliminaries

Lemma C.3. *Let F_X be the cdf of a random variable X and $F_{N,X}$ be the empirical analog. For $\delta \in \left(0, \frac{1}{2}\right)$, if the density f_X exists, is continuous and is bounded away from zero on $\left[F_X^{-1}(\delta), F_X^{-1}(1-\delta)\right]$, then $\sup_{q\in[\delta,1-\delta]}\left|F_{N,X}^{-1}(q) - F_X^{-1}(q)\right| \to 0$ in outer probability.*

Proof. Follows from van der Vaart and Wellner (2000),Example 3.9.21 as a consequence of the continuous mapping theorem. □

Let $v(x,\varepsilon;\theta) = h(x;\theta)+\varepsilon$ be Lipschitz continuous in (x,θ) and let ε have a coninuous density f_{ε}. For a measure m^x on X, let

$$F_{v;\theta}(v) = \int_{-\infty}^{v} F_{\varepsilon}(v - h(x;\theta))\, dm^x$$

and denote the corresponding empirical cdf of a sample size N with $F_{N,v;\theta}$. This is the cdf of the image $m_N^{x,\varepsilon}$ Assume $\theta \in \Theta$, compact.

Lemma C.4. *If $f_{v;\theta}$ is bounded away from 0 for every compact set, then for each $\delta \in (0,1)$, $\sup_{\theta\in\Theta, q\in[\delta,1-\delta]}\left|F_{N,v;\theta}^{-1}(q) - F_{v;\theta}^{-1}(q)\right| \to 0$ in outer probability.*

Proof. The result follows from the fact that the collection of sets, $\{(x,\varepsilon,\theta) : h(x;\theta)+\varepsilon \le v\}$, indexed by (v,θ) are Glivenko-Cantelli since ε admits a density and $h(x,\theta)$ is Lipschitz continuous. Note that $m_N^{x,\varepsilon}$ converges uniformly to $m^x \times m^{\varepsilon}$ over the collection of sets $\{(x,\varepsilon,\theta) : h(x;\theta)+\varepsilon \le v\}$. We now prove continuity of $F_{N,v;\theta}^{-1}(q) - F_{v;\theta}^{-1}(q)$ with respect to $m_N^{x,\varepsilon} - m^x \times m^{\varepsilon}$.

By definition,

$$q = \int_{-\infty}^{F_{v,\theta}^{-1}(q)} dm^{v,\theta} = \int_{-\infty}^{F_{N,v}^{-1}(q)} dm_N^{v,\theta}$$

where $m^{v,\theta}$ and $m_n^{v,\theta}$ are image measures of $m^{x,\varepsilon}$ and $m_N^{x,\varepsilon}$ under $v(x, \varepsilon; \theta)$. Hence, for $\lambda \in (0, 1-\delta)$ and measure $m_N^{x,\varepsilon}$ such that

$$\|m_N^{x,\varepsilon} - m^x \times m^\varepsilon\| < \lambda,$$

we have

$$\begin{aligned} \left| \int_{F_{v,\theta}^{-1}(q)}^{F_{N,v}^{-1}(q)} dm^{v,\theta} \right| &\leq \left| \int_{-\infty}^{F_{N,v;\theta}^{-1}(q)} dm_N^{v,\theta} - \int_{-\infty}^{F_{N,v;\theta}^{-1}(q)} dm^{v,\theta} \right| \\ &\leq \lambda \end{aligned}$$

Since $m^{v;\theta}$ has a density bounded away from 0 on every compact interval uniformly over θ, we have that

$$\left|\left| F_{N,v;\theta}^{-1}(q) - F_{v;\theta}^{-1}(q) \right|\right| \leq \lambda \left(\inf_{\theta \in \Theta, v \in \left[F_v^{-1}(\delta - \lambda), F_v^{-1}(1-\delta+\lambda) \right]} f_{v,\theta} \right)^{-1}.$$

Hence, the result follows by the continuous mapping theorem since $\|m_N^{x,\varepsilon} - m^x \times m^\varepsilon\| \to 0$ by the law of large numbers. □

Lemma C.5. *If Assumption 3.5 is satisfied, then* $\left\|\nabla \tilde{\psi}\right\|_\infty < \infty$.

Proof. Note that

$$\begin{aligned} & \tilde{\psi}(v_1, v_2, u) \\ = & \int \Psi(X_1, X_2, Z)\, dm^{x|v_1} dm^{x|v_2} dm^{z|v} \\ = & \int \Psi(X_1, X_2, Z) \frac{f_\varepsilon(v_1 - h(X_1; \theta_0))}{\int f_\varepsilon(v_1 - h(X_1; \theta_0))\, dm^x} \frac{f_\varepsilon(v_1 - h(X_1; \theta_0))}{\int f_\varepsilon(v_1 - h(X_1; \theta_0))\, dm^x} \frac{f_\eta(u - g(z, \theta_0))}{\int f_\eta(u - g(Z, \theta_0))\, dm^z} dm^x dm^x dm^z \end{aligned}$$

We will only show $\tilde{\psi}(v_1, v_2, u)$ has a bounded derivative in v_1 as the proof for the other two arguments are identical. Note that

$$\begin{aligned} & \frac{\partial}{\partial v} \frac{f_\varepsilon(v - h(x; \theta_0))}{\int f_\varepsilon(v - h(X; \theta_0))\, dm^x} \\ = & \frac{f_\varepsilon'(v - h(x; \theta_0))}{\int f_\varepsilon(v - h(X; \theta_0))\, dm^x} - \frac{f_\varepsilon(v - h(x; \theta_0)) \int f_\varepsilon'(v - h(X; \theta_0))\, dm^x}{\left(\int f_\varepsilon(u - h(X; \theta_0))\, dm^x\right)^2} \quad \text{(C.2)} \end{aligned}$$

If the expression in equation (C.2) is m^x integrable in x, and the terms $\frac{f_\varepsilon(v-h(x;\theta_0))}{\int f_\varepsilon(v-h(X;\theta_0))dm^x}$ and $\frac{f_\eta(u-g(z,\theta_0))}{\int f_\eta(u-g(Z,\theta_0))dm^z}$ are respectively integrable in x' and z, then the Dominated Convergence Theorem implies that the derivative $\frac{\partial}{\partial v_1}\tilde{\psi}(v_1, v_2, u)$ exists and is given by

$$\int \Psi(X_1, X_2, Z) \frac{\partial}{\partial v_1} \frac{f_\varepsilon(v_1 - h(x;\theta_0))}{\int f_\varepsilon(v_1 - h(X;\theta_0)) dm^x} \frac{f_\varepsilon(v_1 - h(X_1;\theta_0))}{\int f_\varepsilon(v_1 - h(X_1;\theta_0)) dm^x} \frac{f_\eta(u - g(z,\theta_0))}{\int f_\eta(u - g(Z,\theta_0)) dm^z} dm^x dm^x dm^z$$

Since $\|\Psi\|_\infty < \infty$, and $\frac{f_\varepsilon(v_1-h(X_1;\theta_0))}{\int f_\varepsilon(v_1-h(X_1;\theta_0))dm^x} \frac{f_\eta(u-g(z,\theta_0))}{\int f_\eta(u-g(Z,\theta_0))dm^z} \leq 1$,

$$\frac{\partial}{\partial v_1}\tilde{\psi}(v_1, v_2, u) \leq \|\Psi\|_\infty \left| \int \left(\frac{f'_\varepsilon(v - h(x;\theta_0))}{\int f_\varepsilon(v - h(X;\theta_0)) dm^X} - \frac{f_\varepsilon(v - h(x;\theta_0)) \int f'_\varepsilon(v - h(X;\theta_0)) dm^X}{\left(\int f_\varepsilon(v - h(X;\theta_0)) dm^X\right)^2} \right) dm^x \right|$$

The result therefore holds if

$$\begin{aligned}
\sup_v \left| \int \left(\frac{f'_\varepsilon(v - h(x;\theta_0))}{\int f_\varepsilon(v - h(X;\theta_0)) dm^X} - \frac{f_\varepsilon(v - h(x;\theta_0)) \int f'_\varepsilon(v - h(X;\theta_0)) dm^X}{\left(\int f_\varepsilon(v - h(X;\theta_0)) dm^X\right)^2} \right) dm^x \right| &< \infty \\
\sup_u \left| \int \left(\frac{f'_\eta(u - g(z;\theta_0))}{\int f_\eta(u - g(Z,\theta_0)) dm^z} - \frac{f_\eta(u - g(z,\theta_0)) \int f'_\eta(u - g(Z,\theta_0)) dm^z}{\left(\int f_\eta(u - g(Z,\theta_0)) dm^z\right)^2} \right) dm^z \right| &< \infty
\end{aligned}$$

Sine the proof is symmetric, we present the argument on for the first inequality. Note that

$$\frac{f'_\varepsilon(v - h(x;\theta_0))}{\int f_\varepsilon(v - h(X;\theta_0)) dm^X} - \frac{f_\varepsilon(v - h(x;\theta_0)) \int f'_\varepsilon(v - h(X;\theta_0)) dm^X}{\left(\int f_\varepsilon(v - h(X;\theta_0)) dm^X\right)^2} = \frac{\partial}{\partial v} \frac{f_\varepsilon(v - h(x;\theta_0))}{f_v}.$$

Hence, we have that

$$\begin{aligned}
& \sup_v \left| \int \left(\frac{f'_\varepsilon(v - h(x;\theta_0))}{\int f_\varepsilon(v - h(X;\theta_0)) dm^X} - \frac{f_\varepsilon(v - h(x;\theta_0)) \int f'_\varepsilon(v - h(X;\theta_0)) dm^X}{\left(\int f_\varepsilon(v - h(X;\theta_0)) dm^X\right)^2} \right) dm^x \right| \\
\leq & \sup_{v,x} \left| \frac{\partial}{\partial v} \frac{f_\varepsilon(v - h(x;\theta_0)) f_X(x)}{f_v(v)} \right| \\
= & \sup_{v,x} \left| \frac{\partial}{\partial v} f_{X|v}(x) \right| < \infty
\end{aligned}$$

by Assumption 3.5 (iv) and (v). □

C.2.2 Proof of Proposition 3.7

Lemma C.6. *If Assumption 3.5 is satisfied, $E(\psi_N | m_N^v, m_J^u) - \psi$ converges in probability to 0 as $N \to \infty$.*

Proof. The quantity $E(\psi_N | m_N^v, m_J^u)$ can be computed from the fact that for all $1 \leq k \leq J$, the k'th most desirable firm is occupied by workers that have the $2k$ and the $(2k-1)$-st most desirable workers. By definition, the conditional expectation of $\Psi(x_1, x_2, z)$ for the k'th desirable job is (or the other way) $\tilde{\psi}\left(F_{N,v}^{-1}\left(\frac{2k-1}{N}\right), F_{N,v}^{-1}\left(\frac{2k}{N}\right), F_{J,u}^{-1}\left(\frac{k}{J}\right)\right)$ where $F_{N,v}$ and $F_{J,u}$ are the cdfs representing the empirical measures m_N^v and m_J^u respectively:

$$\begin{aligned} E\left(\psi_N | m_N^v, m_J^u\right) &= \frac{1}{J}\sum_{k=1}^{J}\tilde{\psi}\left(F_{N,v}^{-1}\left(\frac{2k-1}{N}\right), F_{N,v}^{-1}\left(\frac{2k}{N}\right), F_{J,u}^{-1}\left(\frac{k}{J}\right)\right) \\ &= \frac{1}{2J}\sum_{i=1}^{2J}\tilde{\psi}\left(F_{N,v}^{-1}\left(\frac{i}{N}\right), F_{N,v}^{-1}\left(\frac{i}{N}\right), F_{J,u}^{-1}\left(\frac{i}{N}\right)\right) + R. \end{aligned} \tag{C.3}$$

We will show that

$$\frac{1}{2J}\sum_{i=1}^{2J}\tilde{\psi}\left(F_{N,v}^{-1}\left(\frac{i}{N}\right), F_{N,v}^{-1}\left(\frac{i}{N}\right), F_{J,u}^{-1}\left(\frac{i}{N}\right)\right) - \psi \to 0 \tag{C.4}$$

and that $R \to 0$.

First, consider $|R|$ for large N. By the expression above,

$$\begin{aligned} R = \frac{1}{J}\sum_{k=1}^{J}\Bigg[&\tilde{\psi}\left(F_{N,v}^{-1}\left(\frac{2k-1}{N}\right), F_{N,v}^{-1}\left(\frac{2k}{N}\right), F_{J,u}^{-1}\left(\frac{k}{J}\right)\right) \\ &-\frac{1}{2}\tilde{\psi}\left(F_{N,v}^{-1}\left(\frac{2k-1}{N}\right), F_{N,v}^{-1}\left(\frac{2k-1}{N}\right), F_{J,u}^{-1}\left(\frac{2k-1}{N}\right)\right) \\ &-\frac{1}{2}\tilde{\psi}\left(F_{N,v}^{-1}\left(\frac{2k}{N}\right), F_{N,v}^{-1}\left(\frac{2k}{N}\right), F_{J,u}^{-1}\left(\frac{2k}{N}\right)\right)\Bigg]. \end{aligned}$$

By the triangle ineuqality, the absolute value of the k-th term in the summation is at most

$$\begin{aligned} &\frac{1}{2}\left|\tilde{\psi}\left(F_{N,v}^{-1}\left(\frac{2k-1}{N}\right), F_{N,v}^{-1}\left(\frac{2k}{N}\right), F_{J,u}^{-1}\left(\frac{k}{J}\right)\right) - \tilde{\psi}\left(F_{N,v}^{-1}\left(\frac{2k-1}{N}\right), F_{N,v}^{-1}\left(\frac{2k-1}{N}\right), F_{J,u}^{-1}\left(\frac{2k-1}{N}\right)\right)\right| \\ &+\frac{1}{2}\left|\tilde{\psi}\left(F_{N,v}^{-1}\left(\frac{2k-1}{N}\right), F_{N,v}^{-1}\left(\frac{2k}{N}\right), F_{J,u}^{-1}\left(\frac{k}{J}\right)\right) - \tilde{\psi}\left(F_{N,v}^{-1}\left(\frac{2k}{N}\right), F_{N,v}^{-1}\left(\frac{2k}{N}\right), F_{J,u}^{-1}\left(\frac{2k}{N}\right)\right)\right| \end{aligned}$$

Hence,

$$\begin{aligned} |R| \leq \frac{1}{2J}\sum_{k=1}^{J}\Bigg|&\tilde{\psi}\left(F_{N,v}^{-1}\left(\frac{2k-1}{N}\right), F_{N,v}^{-1}\left(\frac{2k}{N}\right), F_{J,u}^{-1}\left(\frac{2k}{N}\right)\right) \\ &-\tilde{\psi}\left(F_{N,v}^{-1}\left(\frac{2k}{N}\right), F_{N,v}^{-1}\left(\frac{2k}{N}\right), F_{J,u}^{-1}\left(\frac{2k}{N}\right)\right)\Bigg| \\ +\frac{1}{2J}\sum_{k=1}^{J}\Bigg|&\tilde{\psi}\left(F_{N,v}^{-1}\left(\frac{2k-1}{N}\right), F_{N,v}^{-1}\left(\frac{2k}{N}\right), F_{J,u}^{-1}\left(\frac{2k}{N}\right)\right) \\ &-\tilde{\psi}\left(F_{N,v}^{-1}\left(\frac{2k-1}{N}\right), F_{N,v}^{-1}\left(\frac{2k-1}{N}\right), F_{J,u}^{-1}\left(\frac{2k}{N}\right)\right)\Bigg| \end{aligned}$$

For $\delta \in \left(0, \frac{1}{2}\right)$,

$$\begin{aligned}
|R| &\leq \frac{1}{2J} \sum_{J\delta < k < J(1-\delta)} \left| \tilde{\psi}\left(F_{N,v}^{-1}\left(\frac{2k-1}{N}\right), F_{N,v}^{-1}\left(\frac{2k}{N}\right), F_{J,u}^{-1}\left(\frac{2k}{N}\right)\right) \right. \\
&\quad \left. -\tilde{\psi}\left(F_{N,v}^{-1}\left(\frac{2k}{N}\right), F_{N,v}^{-1}\left(\frac{2k}{N}\right), F_{J,u}^{-1}\left(\frac{2k}{N}\right)\right) \right| \\
&\quad +\frac{1}{2J} \sum_{J\delta < k < J(1-\delta)} \left| \tilde{\psi}\left(F_{N,v}^{-1}\left(\frac{2k-1}{N}\right), F_{N,v}^{-1}\left(\frac{2k}{N}\right), F_{J,u}^{-1}\left(\frac{2k}{N}\right)\right) \right. \\
&\quad \left. -\tilde{\psi}\left(F_{N,v}^{-1}\left(\frac{2k-1}{N}\right), F_{N,v}^{-1}\left(\frac{2k-1}{N}\right), F_{J,u}^{-1}\left(\frac{2k}{N}\right)\right) \right| \\
&\quad +4\delta \left\| \tilde{\psi} \right\|_\infty \\
&= R_1 + R_2 + 4\delta \left\| \tilde{\psi} \right\|_\infty \qquad \text{(C.5)}
\end{aligned}$$

By a first order Taylor expansion,

$$R_1 + R_2 \leq \sup_{J\delta < k < J(1-\delta)} \left\| \nabla \tilde{\psi} \right\|_\infty \left[\left| F_{N,v}^{-1}\left(\frac{2k-1}{N}\right) - F_{N,v}^{-1}\left(\frac{2k}{N}\right) \right| + \left| F_{N,v}^{-1}\left(\frac{2k}{N}\right) - F_{N,v}^{-1}\left(\frac{2k}{N}\right) \right| \right] \qquad \text{(C.6)}$$

and since $R_1, R_2 \geq 0$, the sum $R_1 + R_2$ converges in probability to 0 by the uniform convergence of the quantile process of v on $[\delta, 1-\delta]$ (Lemmata C.4, C.9).

Now, consider the difference in equation (C.4). Note that $F_{J,u}$ is constant on each interval $[\frac{k-1}{J}, \frac{k}{J})$ and $F_{N,v}$ is constant on $[\frac{i-1}{N}, \frac{i}{N})$. Hence,

$$\begin{aligned}
&\frac{1}{2J} \sum_{i=1}^{2J} \tilde{\psi}\left(F_{N,v}^{-1}\left(\frac{i}{N}\right), F_{N,v}^{-1}\left(\frac{i}{N}\right), F_{J,u}^{-1}\left(\frac{i}{N}\right)\right) - \psi \\
= &\int_0^1 \tilde{\psi}\left(F_{N,v}^{-1}(q), F_{N,v}^{-1}(q), F_{J,u}^{-1}(q)\right) dq - \int_0^1 \tilde{\psi}\left(F_v^{-1}(q), F_v^{-1}(q), F_u^{-1}(q)\right) dq \\
= &\int_\delta^{1-\delta} \left[\tilde{\psi}\left(F_{N,v}^{-1}(q), F_{N,v}^{-1}(q), F_{J,u}^{-1}(q)\right) - \tilde{\psi}\left(F_v^{-1}(q), F_v^{-1}(q), F_u^{-1}(q)\right)\right] dq \\
&+ \left(\int_0^\delta + \int_{1-\delta}^1\right) \left[\tilde{\psi}\left(F_{N,v}^{-1}(q), F_{N,v}^{-1}(q), F_{J,u}^{-1}(q)\right) - \tilde{\psi}\left(F_v^{-1}(q), F_v^{-1}(q), F_u^{-1}(q)\right)\right] dq \\
= &\; T_1 + T_2 \qquad \text{(C.7)}
\end{aligned}$$

where $\delta \in \left(0, \frac{1}{2}\right)$.

We now bound for T_1 and T_2 in terms of $\delta > 0$ and then minimize over δ. Since $\|\psi\|_\infty \leq \|\Psi\|_\infty < \infty$, $|T_2| \leq 4\delta \left\|\tilde{\psi}\right\|_\infty$. To bound T_1, note that

$$
\begin{aligned}
T_1 = & \left| \int_\delta^{1-\delta} \left[\tilde{\psi}\left(F_{N,v}^{-1}(q), F_{N,v}^{-1}(q), F_{J,u}^{-1}(q)\right) - \tilde{\psi}\left(F_v^{-1}(q), F_v^{-1}(q), F_u^{-1}(q)\right) \right] dq \right| \\
\leq & \int_\delta^{1-\delta} \left| \tilde{\psi}\left(F_{N,v}^{-1}(q), F_{N,v}^{-1}(q), F_{J,u}^{-1}(q)\right) - \tilde{\psi}\left(F_v^{-1}(q), F_v^{-1}(q), F_u^{-1}(q)\right) \right| dq
\end{aligned}
$$

and for all $q \in [\delta, 1-\delta]$, we have a bound via a pointwise Taylor expansion,

$$
\begin{aligned}
& \left| \tilde{\psi}\left(F_{N,v}^{-1}(q), F_{N,v}^{-1}(q), F_{J,u}^{-1}(q)\right) - \tilde{\psi}\left(F_v^{-1}(q), F_v^{-1}(q), F_u^{-1}(q)\right) \right| \\
\leq & \left\|\nabla\tilde{\psi}\right\|_\infty \sup_{q\in[\delta,1-\delta]} \left\| \left(F_{N,v}^{-1}(q), F_{N,v}^{-1}(q), F_{J,u}^{-1}(q)\right) - \left(F_v^{-1}(q), F_v^{-1}(q), F_u^{-1}(q)\right) \right\|_\infty
\end{aligned} \tag{C.8}
$$

Note that Lemma C.5 implies that $\left\|\nabla\tilde{\psi}\right\|_\infty$ exists. Combining equations (C.3) - (C.8), and bounds on $R_1 + R_2$ and T_2, we have that

$$
\begin{aligned}
& \left|E\left(\psi_N | m_N^v, m_J^u\right) - \psi\right| \\
\leq & \left\|\nabla\tilde{\psi}\right\|_\infty \sup_{q\in[\delta,1-\delta]} \left| \left(F_{N,v}^{-1}(q), F_{N,v}^{-1}(q), F_{J,u}^{-1}(q)\right) - \left(F_v^{-1}(q), F_v^{-1}(q), F_u^{-1}(q)\right) \right| + 8\delta \|m\|_\infty
\end{aligned}
$$

where $8\delta \|m\|_\infty$ is the contribution from $R_3 + T_2$.

We now show that $\left|E\left(\psi_N | m_N^v, m_J^u\right) - \psi\right| \to 0$ in probability as $N \to \infty$. Fix $\epsilon > 0$ and choose $\delta = \frac{\epsilon}{16\|m\|_\infty}$. Since

$$
\sup_{q\in[\delta,1-\delta]} \left| \left(F_{N,v}^{-1}(q), F_{N,v}^{-1}(q), F_{J,u}^{-1}(q)\right) - \left(F_v^{-1}(q), F_v^{-1}(q), F_u^{-1}(q)\right) \right|
$$

converges in probability to 0, for sufficiently large N we have by Lemma C.4,

$$
P\left(\left\|\nabla\tilde{\psi}\right\|_\infty \sup_{Q'\epsilon[\delta,1-\delta]} \left| \left(F_{N,v}^{-1}(q), F_{N,v}^{-1}(q), F_{J,u}^{-1}(q)\right) - \left(F_v^{-1}(q), F_v^{-1}(q), F_u^{-1}(q)\right) \right| > \frac{\epsilon}{2} \right) < \epsilon.
$$

This implies $P\left(\left|E\left(\psi_N | m_N^v, m_J^u\right) - \psi\right| > \epsilon\right) < \epsilon$, proving the desired convergence in probability to 0. □

Lemma C.7. *$\psi_N - E\left(\psi_N | m_N^v, m_J^u\right)$ converges in probability to 0 if $\|\Psi\|_\infty < \infty$.*

Proof. Let $v^{(k)}$ and $u^{(k)}$ be k'th order statistics of worker and firm desirability and let $X^{(k)}$ and $Z^{(k)}$ be the corresponding observations drawn from $m^{x|v^{(k)}}$ and $m^{z|u^{(k)}}$ respectively.

Note that the second moment of the function

$$\psi_N - E\left(\psi_N | m_N^u, m_N^v\right) = \frac{1}{J}\left(\sum_{k=1}^{J} \Psi\left(X^{(2k-1)}, X^{(2k)}, Z^{(i)}\right) - \tilde{\psi}\left(v^{(2k-1)}, v^{(2k)}, u^{(k)}\right)\right).$$

conditional on $\left(m_N^v, m_J^u\right)$ is

$$\begin{aligned} & \frac{1}{J^2} E\left(\sum_{i=1}^{J} \Psi\left(X^{(2k-1)}, X^{(2k)}, Z^{(k)}\right) - \tilde{\psi}\left(v^{(2k-1)}, v^{(2k)}, u^{(k)}\right) | m_N^v, m_J^u\right)^2 \\ = & \frac{1}{J^2} \sum_{i=1}^{J} E\left(\Psi\left(X^{(2k-1)}, X^{(2k)}, Z^{(k)}\right) - \tilde{\psi}\left(v^{(2k-1)}, v^{(2k)}, u^{(k)}\right) | m_N^v, m_J^u\right) \\ \leq & \frac{1}{J}\left\|\sigma^2\right\|_{\infty}, \end{aligned}$$

where $\left\|\sigma^2\right\|_{\infty}$ is the supremum of the function

$$\sigma^2\left(v_1, v_2, u\right) = Var\left(\Psi\left(X_1, X_2, Z\right) | h\left(X_1; \theta_0\right) + \varepsilon_1 = v_1, h\left(X_2; \theta_0\right) + \varepsilon_2 = v_2, g\left(Z; \theta_0\right) + \eta = u\right).$$

The quantity $\left\|\sigma^2\right\|_{\infty}$ is well defined and finite since $\Psi\left(x_1, x_2, z\right)$ is bounded.

However, since $\psi_N - E\left(\psi_N | m_N^v, m_J^u\right)$ is by definition mean zero, it follows that the unconditional variance of $\psi_N - E\left(\psi_N | m_N^v, m_J^u\right)$ is bounded above by $\frac{1}{J}\left\|\sigma^2\right\|_{\infty}$, by the law of total variance.. This proves $\sqrt{N}\left(\psi_N - E\left(\psi_N | m_N^v, m_J^u\right)\right) = O_p(1)$ and thus, by Chevyshev's inequality, $\psi_N - E\left(\psi_N | m_N^v, m_J^u\right) = o_p(1)$. □

Proposition 3.7 *Let ψ^k and ψ_N^k denote the k-th dimensions of ψ and ψ_N respectively. If Assumption 3.5 is satisfied, then for each $k \in \{1, \ldots, K_\Psi\}$, $\psi_N^k - \psi^k$ converges in probability to* 0.

Proof. The proof is identical for each component k. For notational simplicity we drop the index k. Since $\psi_N - \psi = \left(\psi_N - E\left(\psi_N | m_N^v, m_J^u\right)\right) + \left(E\left(\psi_N | m_N^v, m_J^u\right) - \psi\right)$ is the sum of two terms that converge in probability to 0, this follows directly from Slutsky's theorem. □

C.2.3 Proof of Proposition 3.8

For ease of notation, define the quantities

$$\begin{aligned}
\varepsilon(q,x;\theta) &= F_{v;\theta}^{-1}(q)-h(x;\theta)\\
\varepsilon_N(q,x;\theta) &= F_{N,u;\theta}^{-1}(q)-h(x;\theta)\\
\eta(q,z;\theta) &= F_{u;\theta}^{-1}(q)-g(z;\theta)\\
\eta_N(q,z;\theta) &= F_{N,u;\theta}^{-1}(q)-g(z;\theta).
\end{aligned}$$

We first prove two preliminary results

Lemma C.8. *If Assumption 3.5(i) - (iii) are satisfied, then for each* $\delta\in\left(0,\frac{1}{2}\right)$*, the quantity*

$$T_{1,\delta}=\sup_{\theta}\int_{\delta}^{1-\delta}\left|t_1(q;m^x,m^z)-t_{N,1}(q;m_N^x,m_N^z)\right|dq$$

where

$$\begin{aligned}
t_1(q;m^x,m^z,\theta) &= \int\Psi(X_1,X_2,Z)\,f_{\varepsilon}(\varepsilon(q,X_1;\theta))\,f_{\varepsilon}(\varepsilon(q,X_2;\theta))\,f_{\eta}(\eta(q,Z;\theta))\,dm^x dm^x dm^z\\
t_{N,1}(q;m^x,m^z,\theta) &= \int\Psi(X_1,X_2,Z)\,f_{\varepsilon}(\varepsilon_N(q,X_1;\theta))\,f_{\varepsilon}(\varepsilon_N(q,X_2;\theta))\,f_{\eta}(\eta_N(q,Z;\theta))\,dm^x dm^x dm^z
\end{aligned}$$

converges in outer probability to 0*.*

Proof. We first split $T_{1,\delta}$ into two terms R_1 and R_2 as follows:

$$\begin{aligned}
&\sup_{\theta}\int_{\delta}^{1-\delta}\left|t_1(q;m^x,m^z,\theta)-t_{N,1}(q;m_N^x,m_N^z,\theta)\right|dq\\
\leq\ &\sup_{\theta}\int_{\delta}^{1-\delta}\left|t_1(q;m^x,m^z,\theta)-t_{N,1}(q;m^x,m^z,\theta)\right|dq\\
&+\sup_{\theta}\int_{\delta}^{1-\delta}\left|t_{N,1}(q;m^x,m^z,\theta)-t_{N,1}(q;m_N^x,m_N^z,\theta)\right|dq\\
=\ &R_1+R_2
\end{aligned}$$

R_2 converges to 0 by a Glivenko-Cantelli theorem (result 3).

To bound R_1, note that

$$
\begin{aligned}
& \sup_{\theta} \int_{\delta}^{1-\delta} \left| t_1 \left(q; m^x, m^z \right) - t_{N,1} \left(q; m^x, m^z \right) \right| dq \\
= & \int_{\delta}^{1-\delta} \left| \int \Psi \left(X_1, X_2, Z \right) \left[f_{\varepsilon} \left(\varepsilon \left(q, X_1; \theta \right) \right) f_{\varepsilon} \left(\varepsilon \left(q, X_2; \theta \right) \right) f_{\eta} \left(\eta \left(q, Z; \theta \right) \right) \right. \right. \\
& \left. \left. - f_{\varepsilon} \left(\varepsilon_N \left(q, X_1; \theta \right) \right) f_{\varepsilon} \left(\varepsilon_N \left(q, X_2; \theta \right) \right) f_{\eta} \left(\eta_N \left(q, Z; \theta \right) \right) \right] dm^x dm^x dm^z dq \right| \\
\leq & \left\| \Psi \right\|_{\infty} \sup_{\theta, q \epsilon [\delta, 1-\delta]} \left| f_{\varepsilon} \left(\varepsilon \left(q, X_1; \theta \right) \right) f_{\varepsilon} \left(\varepsilon \left(q, X_2; \theta \right) \right) f_{\eta} \left(\eta \left(q, Z; \theta \right) \right) - \right. \\
& \left. f_{\varepsilon} \left(\varepsilon_N \left(q, X_1; \theta \right) \right) f_{\varepsilon} \left(\varepsilon_N \left(q, X_2; \theta \right) \right) f_{\eta} \left(\eta_N \left(q, Z; \theta \right) \right) \right| .
\end{aligned}
$$

We now show that

$$
\sup_{\theta, q \epsilon [\delta, 1-\delta]} \left| f_{\varepsilon} \left(\varepsilon \left(q, X_1; \theta \right) \right) f_{\varepsilon} \left(\varepsilon \left(q, X_2; \theta \right) \right) f_{\eta} \left(\eta \left(q, Z; \theta \right) \right) - f_{\varepsilon} \left(\varepsilon_N \left(q, X_1; \theta \right) \right) f_{\varepsilon} \left(\varepsilon_N \left(q, X_2; \theta \right) \right) f_{\eta} \left(\eta_N \left(q, Z; \theta \right) \right) \right|
$$

converges to 0 in outer probability. First rewrite the difference

$$
\begin{aligned}
& f_{\varepsilon} \left(\varepsilon \left(q, x_1; \theta \right) \right) f_{\varepsilon} \left(\varepsilon \left(q, x_2; \theta \right) \right) f_{\eta} \left(\eta \left(q, z; \theta \right) \right) - f_{\varepsilon} \left(\varepsilon_N \left(q, x_1; \theta \right) \right) f_{\varepsilon} \left(\varepsilon_N \left(q, x_2; \theta \right) \right) f_{\eta} \left(\eta_N \left(q, z; \theta \right) \right) \\
& = f_{\varepsilon} \left(\varepsilon \left(q, x_1; \theta \right) \right) f_{\varepsilon} \left(\varepsilon \left(q, x_2; \theta \right) \right) \left[f_{\eta} \left(\eta \left(q, z; \theta \right) \right) - f_{\eta} \left(\eta_N \left(q, z; \theta \right) \right) \right] \\
& + f_{\eta} \left(\eta \left(q, z; \theta \right) \right) \left[f_{\varepsilon} \left(\varepsilon \left(q, x_1; \theta \right) \right) f_{\varepsilon} \left(\varepsilon \left(q, x_2; \theta \right) \right) - f_{\varepsilon} \left(\varepsilon_N \left(q, x_1; \theta \right) \right) f_{\varepsilon} \left(\varepsilon_N \left(q, x_2; \theta \right) \right) \right] \\
& + \left[f_{\eta} \left(\eta_N \left(q, z; \theta \right) \right) - f_{\eta} \left(\eta \left(q, z; \theta \right) \right) \right] \left[f_{\varepsilon} \left(\varepsilon \left(q, x_1; \theta \right) \right) f_{\varepsilon} \left(\varepsilon \left(q, x_2; \theta \right) \right) - f_{\varepsilon} \left(\varepsilon_N \left(q, x_1; \theta \right) \right) f_{\varepsilon} \left(\varepsilon_N \left(q, x_2; \theta \right) \right) \right] .
\end{aligned}
$$

Hence, by the triangle inequality,

$$
\begin{aligned}
& \left| f_{\varepsilon} \left(\varepsilon \left(q, x_1; \theta \right) \right) f_{\varepsilon} \left(\varepsilon \left(q, x_2; \theta \right) \right) f_{\eta} \left(\eta \left(q, z; \theta \right) \right) - f_{\varepsilon} \left(\varepsilon_N \left(q, x_1; \theta \right) \right) f_{\varepsilon} \left(\varepsilon_N \left(q, x_2; \theta \right) \right) f_{\eta} \left(\eta_N \left(q, z; \theta \right) \right) \right| \\
& \leq f_{\varepsilon} \left(\varepsilon \left(q, x_1; \theta \right) \right) f_{\varepsilon} \left(\varepsilon \left(q, x_2; \theta \right) \right) \left| f_{\eta} \left(\eta \left(q, z; \theta \right) \right) - f_{\eta} \left(\eta_N \left(q, z; \theta \right) \right) \right| \\
& + f_{\eta} \left(\eta \left(q, z; \theta \right) \right) \left| f_{\varepsilon} \left(\varepsilon \left(q, x_1; \theta \right) \right) f_{\varepsilon} \left(\varepsilon \left(q, x_2; \theta \right) \right) - f_{\varepsilon} \left(\varepsilon_N \left(q, x_1; \theta \right) \right) f_{\varepsilon} \left(\varepsilon_N \left(q, x_2; \theta \right) \right) \right| \\
& + \left| f_{\eta} \left(\eta_N \left(q, z; \theta \right) \right) - f_{\eta} \left(\eta \left(q, z; \theta \right) \right) \right| \left| f_{\varepsilon} \left(\varepsilon \left(q, x_1; \theta \right) \right) f_{\varepsilon} \left(\varepsilon \left(q, x_2; \theta \right) \right) - f_{\varepsilon} \left(\varepsilon_N \left(q, x_1; \theta \right) \right) f_{\varepsilon} \left(\varepsilon_N \left(q, x_2; \theta \right) \right) \right|
\end{aligned}
$$

Further, since f_{ε} and f_{η} are bounded,

$$
\left| f_{\varepsilon} \left(\varepsilon \left(q, x_1; \theta \right) \right) f_{\varepsilon} \left(\varepsilon \left(q, x_2; \theta \right) \right) f_{\eta} \left(\eta \left(q, z; \theta \right) \right) - f_{\varepsilon} \left(\varepsilon_N \left(q, x_1; \theta \right) \right) f_{\varepsilon} \left(\varepsilon_N \left(q, x_2; \theta \right) \right) f_{\eta} \left(\eta_N \left(q, z; \theta \right) \right) \right|
$$

is bounded by

$$
\begin{aligned}
&\|f_\varepsilon\|_\infty^2 \left|f_\eta\left(\eta\left(q,z;\theta\right)\right)-f_\eta\left(\eta_N\left(q,z;\theta\right)\right)\right| \\
&+\|f_\eta\|_\infty \left|f_\varepsilon\left(\varepsilon\left(q,x_1;\theta\right)\right)f_\varepsilon\left(\varepsilon\left(q,x_2;\theta\right)\right)-f_\varepsilon\left(\varepsilon_N\left(q,x_1;\theta\right)\right)f_\varepsilon\left(\varepsilon_N\left(q,x_2;\theta\right)\right)\right| \\
&+2\|f_\eta\|_\infty \left|f_\varepsilon\left(\varepsilon\left(q,x_1;\theta\right)\right)f_\varepsilon\left(\varepsilon\left(q,x_2;\theta\right)\right)-f_\varepsilon\left(\varepsilon_N\left(q,x_1;\theta\right)\right)f_\varepsilon\left(\varepsilon_N\left(q,x_2;\theta\right)\right)\right| \\
&\ \le \|f_\varepsilon\|_\infty^2 \left\|f_\eta'\right\|_\infty \left|\eta\left(q,z;\theta\right)-\eta_N\left(q,z;\theta\right)\right| \\
&+3\|f_\eta\|_\infty \|f_\varepsilon\|_\infty \|f_\varepsilon'\|_\infty \left|\varepsilon\left(q,x_1;\theta\right)-\varepsilon_N\left(q,x_1;\theta\right)\right|
\end{aligned}
$$

where the last inequality follows from a Taylor expansion. Hence, we have that

$$
\begin{aligned}
&\|\Psi\|_\infty \sup_{\theta,q\epsilon[\delta,1-\delta]} \left|f_\varepsilon\left(\varepsilon\left(q,x_1;\theta\right)\right)f_\varepsilon\left(\varepsilon\left(q,x_2;\theta\right)\right)f_\eta\left(\eta\left(q,z;\theta\right)\right)\right. \\
&\left.-\ f_\varepsilon\left(\varepsilon_N\left(q,x_1;\theta\right)\right)f_\varepsilon\left(\varepsilon_N\left(q,x_2;\theta\right)\right)f_\eta\left(\eta_N\left(q,z;\theta\right)\right)\right| \\
\le\ &\|\Psi\|_\infty \|f_\varepsilon\|^2 \left\|f_\eta'\right\| \sup_{\theta\in\Theta}\left|\eta\left(q,z;\theta\right)-\eta_N\left(q,z;\theta\right)\right| \\
&+3\|\Psi\|_\infty \|f_\eta\| \|f_\varepsilon\|_\infty \|f_\varepsilon'\| \sup_{\theta\in\Theta}\left|\varepsilon\left(q,x_1;\theta\right)-\varepsilon_N\left(q,x_1;\theta\right)\right|.
\end{aligned}
$$

By Lemma C.4, $\sup_{\theta\in\Theta,q\in[\delta,1-\delta]}\left|\varepsilon\left(q,z;\theta\right)-\varepsilon_N\left(q,z;\theta\right)\right|$ and $\sup_{\theta\in\Theta,q\in[\delta,1-\delta]}\left|\eta\left(q,z;\theta\right)-\eta_N\left(q,z;\theta\right)\right|$ converge in outer probability to 0. Hence, $|R_1|$ is bounded by a function that converges in outter probability to 0.

Since $T_{1,\delta}$ is bounded above by the sum of elements which converge in outer probability to 0, it does so as well. □

Lemma C.9. *If Assumptions 3.5(ii) and (iii) are satisfied, then for every $\delta\in\left(0,\frac{1}{2}\right)$ and $q\in[\delta,1-\delta]$, the quantities*

$$
\int f_\varepsilon\left(\varepsilon\left(q,X_1;\theta\right)\right)dm^x \text{ and } \int f_\eta\left(\eta\left(q,Z;\theta\right)\right)dm^z
$$

are bounded away from 0, uniformly in θ. In particular,

$$
\int f_\varepsilon\left(\varepsilon\left(q,X_1;\theta\right)\right)f_\varepsilon\left(\varepsilon\left(q,X_2;\theta\right)\right)f_\eta\left(\eta\left(q,Z;\theta\right)\right)dm^x dm^x dm^z
$$

is also bounded away from zero.

Proof. Note that

$$
\begin{aligned}
&\int f_\varepsilon\left(\varepsilon\left(q,X_1;\theta\right)\right)f_\varepsilon\left(\varepsilon\left(q,X_2;\theta\right)\right)f_\eta\left(\eta\left(q,Z;\theta\right)\right)dm^x dm^x dm^z \\
&= f_{v;\theta}\left(F_{v;\theta}^{-1}\left(q\right)\right)f_{v;\theta}\left(F_{v;\theta}^{-1}\left(q\right)\right)f_{u;\theta}\left(F_{u;\theta}^{-1}\left(q\right)\right)
\end{aligned}
$$

where $f_{u;\theta}\left(F_{u;\theta}^{-1}(q)\right)$ and $f_{v;\theta}\left(F_{v;\theta}^{-1}(q)\right)$ are the densities of u and v at their q-th quantiles respectively. Assumptions 3.5(ii) and (iii) require that f_ε and f_η are continuous and strictly positive, hence they bounded away from zero on any compact set. Consequently, $f_{u;\theta}$ and $f_{v;\theta}$ are also bounded away from zero on any compact set. Since $F_{u;\theta}^{-1}$ is jointly continuous, the image $F_{u;\theta}^{-1}([\delta, 1-\delta])$ is compact and $f_{v;\theta}\left(F_v^{-1}(q)\right) f_{v;\theta}\left(F_v^{-1}(q)\right) f_{u;\theta}\left(F_u^{-1}(q)\right)$ is bounded away from 0 for all $q \in [\delta, 1-\delta]$. □

Lemma C.10. *If Assumptions 3.5(ii) and (iii) are satisfied, the quantity*

$$\begin{aligned} T_{2,\delta} &= \sup_{\theta\in\Theta}\int_\delta^{1-\delta}\left|\frac{1}{\int f_\varepsilon\left(\varepsilon\left(q,X_1;\theta\right)\right) f_\varepsilon\left(\varepsilon\left(q,X_2;\theta\right)\right) f_\eta\left(\eta\left(q,Z;\theta\right)\right) dm^x dm^x dm^z}\right. \\ &\quad \left. -\frac{1}{\int f_\varepsilon\left(\varepsilon_N\left(q,X_1;\theta\right)\right) f_\varepsilon\left(\varepsilon_N\left(q,X_2;\theta\right)\right) f_\eta\left(\eta\left(q,Z;\theta\right)\right) dm_N^x dm_N^x dm_J^z}\right| dq \end{aligned}$$

converges in outer probability to 0.

Proof. As in proof of Lemma C.8,

$$\begin{aligned} L_N &= \sup_{\theta, q\in[\delta,1-\delta]}\left|\int f_\varepsilon\left(\varepsilon\left(q,X_1;\theta\right)\right) f_\varepsilon\left(\varepsilon\left(q,X_2;\theta\right)\right) f_\eta\left(\eta\left(q,Z;\theta\right)\right) dm^x dm^x dm^z -\right. \\ &\quad \left.\int f_\varepsilon\left(\varepsilon_N\left(q,X_1;\theta\right)\right) f_\varepsilon\left(\varepsilon_N\left(q,X_2;\theta\right)\right) f_\eta\left(\eta_N\left(q,Z;\theta\right)\right)\right| dm_N^x dm_N^x dm_J^z \end{aligned}$$

converges in outer probability to 0.

Since $\int f_\varepsilon\left(\varepsilon\left(q,X_1;\theta\right)\right) f_\varepsilon\left(\varepsilon\left(q,X_2;\theta\right)\right) f_\eta\left(\eta\left(q,Z;\theta\right)\right) dm^x dm^x dm^z$ is bounded away from 0 over $q \in [\delta, 1-\delta]$ and all θ (Lemma C.9), by a tailor expansion of the function $1/x$, for every $\epsilon > 0$ there exists $\delta > 0$ such that

$$\begin{aligned} &L_N < \delta \\ &\quad \Rightarrow \sup_{\theta, q\epsilon[\delta,1-\delta]}\left|\frac{1}{\int f_\varepsilon\left(\varepsilon\left(q,X_1;\theta\right)\right) f_\varepsilon\left(\varepsilon\left(q,X_2;\theta\right)\right) f_\eta\left(\eta\left(q,Z;\theta\right)\right) dm^x dm^x dm^z} -\right. \\ &\quad \left.\frac{1}{\int f_\varepsilon\left(\varepsilon_N\left(q,X_1;\theta\right)\right) f_\varepsilon\left(\varepsilon_N\left(q,X_2;\theta\right)\right) f_\eta\left(\eta_N\left(q,Z;\theta\right)\right) dm_N^x dm_N^x dm_J^z}\right| < \epsilon. \end{aligned}$$

Therefore, $T_{2,\delta}$ converges in outer probability to 0. □

Proposition 3.8 *Let $\psi^k(\theta)$ and $\psi_N^k(\theta; m_N^x, m_J^z)$ denote the k-th dimensions of $\psi(\theta)$ and $\psi_N(\theta; m_N^x, m_J^z)$ respectively. If Assumptions 3.5(i) - (iii) are satisfied, then for each $k \in \{1, \ldots, K_\Psi\}$, $\left|\psi^k(\theta) - \psi_N^k(\theta; m_N^x, m_J^z)\right|$ converges in outer probability to 0 uniformly in θ.*

Proof. The proof is identical for each dimention of $\psi^k(\theta)$ and $\psi_N^k(\theta; m_N^x, m_J^z)$. For notational simplicity, we drop the index for the dimension k.

Rewrite

$$\psi(\theta) - \psi_N(\theta; m_N^x, m_N^z) = \int_0^1 t(q, \theta; m_N^x, m_N^z)\, dq$$

where

$$\begin{aligned}
&t(q, \theta; m_N^x, m_N^z) \\
&= \frac{\int \Psi(X_1, X_2, Z) f_\varepsilon(\varepsilon(q, X_1; \theta)) f_\varepsilon(\varepsilon(q, X_2; \theta)) f_\eta(\eta(q, Z; \theta))\, dm^x dm^x dm^z}{\int f_\varepsilon(\varepsilon(q, X_1; \theta)) f_\varepsilon(\varepsilon(q, X_2; \theta)) f_\eta(\eta(q, Z; \theta))\, dm^x dm^x dm^z} \\
&- \frac{\int \Psi(X_1, X_2, Z) f_\varepsilon(\varepsilon(q, X_1; \theta)) f_\varepsilon(\varepsilon(q, X_2; \theta)) f_\eta(\eta(q, Z; \theta))\, dm_N^x dm_N^x dm_J^z}{\int f_\varepsilon(\varepsilon(q, X_1; \theta)) f_\varepsilon(\varepsilon(q, X_2; \theta)) f_\eta(\eta(q, Z; \theta))\, dm_N^x dm_N^x dm_J^z}.
\end{aligned}$$

For $\delta \in \left(0, \frac{1}{2}\right)$, we have that

$$\begin{aligned}
\psi(\theta) - \psi_N(\theta; m_N^x, m_N^z) &= \int_\delta^{1-\delta} t(q, \theta; m_N^x, m_N^z)\, dq + \left(\int_0^\delta + \int_{1-\delta}^1\right) t(q, \theta; m_N^x, m_N^z)\, dq \\
&= R_1 + R_2 + \left(\int_0^\delta + \int_{1-\delta}^1\right) t(q, \theta; m_N^x, m_N^z)\, dq \qquad \text{(C.9)}
\end{aligned}$$

where R_1 is

$$\begin{aligned}
&\int_\delta^{1-\delta} \frac{\int \Psi(X_1, X_2, Z) f_\varepsilon(\varepsilon(q, X_1; \theta)) f_\varepsilon(\varepsilon(q, X_2; \theta)) f_\eta(\eta(q, Z; \theta))\, dm^x dm^x dm^z}{\int f_\varepsilon(\varepsilon(q, X_1; \theta)) f_\varepsilon(\varepsilon(q, X_2; \theta)) f_\eta(\eta(q, Z; \theta))\, dm^x dm^x dm^z} dq \\
&- \int_\delta^{1-\delta} \frac{\int \Psi(X_1, X_2, Z) f_\varepsilon(\varepsilon(q, X_1; \theta)) f_\varepsilon(\varepsilon(q, X_2; \theta)) f_\eta(\eta(q, Z; \theta))\, dm_N^x dm_N^x dm_J^z}{\int f_\varepsilon(\varepsilon(q, X_1; \theta)) f_\varepsilon(\varepsilon(q, X_2; \theta)) f_\eta(\eta(q, Z; \theta))\, dm^x dm^x dm^z} dq
\end{aligned}$$

and R_2 is

$$\begin{aligned}
&\int_\delta^{1-\delta} \frac{\int \Psi(X_1, X_2, Z) f_\varepsilon(\varepsilon_N(q, X_1; \theta)) f_\varepsilon(\varepsilon_N(q, X_2; \theta)) f_\eta(\eta_N(q, Z; \theta))\, dm_N^x dm_N^x dm_J^z}{\int f_\varepsilon(\varepsilon(q, X_1; \theta)) f_\varepsilon(\varepsilon(q, X_2; \theta)) f_\eta(\eta(q, Z; \theta))\, dm^x dm^x dm^z} dq \\
&- \int_\delta^{1-\delta} \frac{\int \Psi(X_1, X_2, Z) f_\varepsilon(\varepsilon_N(q, X_1; \theta)) f_\varepsilon(\varepsilon_N(q, X_2; \theta)) f_\eta(\eta_N(q, Z; \theta))\, dm_N^x dm_N^x dm_J^z}{\int f_\varepsilon(\varepsilon_N(q, X_1; \theta)) f_\varepsilon(\varepsilon_N(q, X_2; \theta)) f_\eta(\eta_N(q, Z; \theta))\, dm_N^x dm_N^x dm_J^z} dq.
\end{aligned}$$

Hence, by the triangle inequality,

$$|\psi(\theta) - \psi_N(\theta; m_N^x, m_N^z)| \le |R_1| + |R_2| + \left|\left(\int_0^\delta + \int_{1-\delta}^1\right) t(q, \theta; m_N^x, m_N^z)\, dq\right|. \qquad \text{(C.10)}$$

We now bound each of the terms. Note that the third term is bounded since M is:

$$\left|\left(\int_0^\delta + \int_{1-\delta}^1\right) t(q, \theta; m_N^x, m_N^z)\, dq\right| \le 4\delta \|M\|_\infty. \qquad \text{(C.11)}$$

The term R_1 is bounded in terms of $T_{1,\delta}$ defined in Lemma C.8:

$$\begin{aligned}
|R_1| &\leq \sup_{\theta, q \in [\delta, 1-\delta]} \frac{1}{\int f_\varepsilon\left(\varepsilon\left(q, X_1; \theta\right)\right) f_\varepsilon\left(\varepsilon\left(q, X_2; \theta\right)\right) f_\eta\left(\eta\left(q, Z; \theta\right)\right) dm^x dm^x dm^z} \\
&\quad \times \int_\delta^{1-\delta} \left| \int \Psi\left(X_1, X_2, Z\right) f_\varepsilon\left(\varepsilon\left(q, X_1; \theta\right)\right) f_\varepsilon\left(\varepsilon\left(q, X_2; \theta\right)\right) f_\eta\left(\eta\left(q, Z; \theta\right)\right) dm^x dm^x dm^z \right. \\
&\quad \left. - \int \Psi\left(X_1, X_2, Z\right) f_\varepsilon\left(\varepsilon_N\left(q, X_1; \theta\right)\right) f_\varepsilon\left(\varepsilon_N\left(q, X_2; \theta\right)\right) f_\eta\left(\eta_N\left(q, Z; \theta\right)\right) dm_N^x dm_J^z \right| dq \\
&= \sup_{\theta, q \varepsilon [\delta, 1-\delta]} \frac{1}{\int f_\varepsilon\left(\varepsilon\left(q, X_1; \theta\right)\right) f_\varepsilon\left(\varepsilon\left(q, X_2; \theta\right)\right) f_\eta\left(\eta\left(q, Z; \theta\right)\right) dm^x dm^x dm^z} T_{1,\delta} \qquad \text{(C.12)}
\end{aligned}$$

and finally, the term R_2 is bounded in terms of $T_{2,\delta}$ defined in Lemma C.10:

$$\begin{aligned}
|R_2| &\leq \|\Psi\|_\infty \int_\delta^{1-\delta} \left| \frac{1}{\int f_\varepsilon\left(\varepsilon\left(q, X_1; \theta\right)\right) f_\varepsilon\left(\varepsilon\left(q, X_2; \theta\right)\right) f_\eta\left(\eta\left(q, Z; \theta\right)\right) dm^x dm^x dm^z} \right. \\
&\quad \left. - \frac{1}{\int f_\varepsilon\left(\varepsilon_N\left(q, X_1; \theta\right)\right) f_\varepsilon\left(\varepsilon_N\left(q, X_2; \theta\right)\right) f_\eta\left(\eta_N\left(q, Z; \theta\right)\right) dm_N^x dm_N^x dm_J^z} \right| dq \\
&= \|\Psi\|_\infty T_{2,\delta} \qquad \text{(C.13)}
\end{aligned}$$

For any ϵ, δ can be chosen to make $4\delta \|\Psi\|_\infty \leq \epsilon$ without reference to θ. Hence, we need to show that the terms $|R_1|$ and $|R_2|$ can also be made small enough, uniformly in θ. By equations (C.10), (C.11), (C.12) and (C.13),

$$\begin{aligned}
&\sup_{\theta \in \Theta} \left|\psi\left(\theta\right) - \psi_N\left(\theta; m_N^x, m_N^z\right)\right| \\
\leq &\left(\sup_{\theta, q \in [\delta, 1-\delta]} \frac{1}{\int f_\varepsilon\left(\varepsilon\left(q, X_1; \theta\right)\right) f_\varepsilon\left(\varepsilon\left(q, X_2; \theta\right)\right) f_\eta\left(\eta\left(q, Z; \theta\right)\right) dm^x dm^x dm^z} \right) T_{1,\delta} \\
&+ \|\Psi\|_\infty T_{2,\delta} + 4\delta \|\Psi\|_\infty,
\end{aligned}$$

where $V_{1,\delta}$ and $V_{2,\delta}$ each converge in outer probability to 0.

The term

$$\sup_{\theta, q \in [\delta, 1-\delta]} \frac{1}{\int f_\varepsilon\left(\varepsilon\left(q, X_1; \theta\right)\right) f_\varepsilon\left(\varepsilon\left(q, X_2; \theta\right)\right) f_\eta\left(\eta\left(q, Z; \theta\right)\right) dm^x dm^x dm^z}$$

is bounded since $\int f_\varepsilon\left(\varepsilon\left(q, X_1; \theta\right)\right) f_\varepsilon\left(\varepsilon\left(q, X_2; \theta\right)\right) f_\eta\left(\eta\left(q, Z; \theta\right)\right) dm^x dm^x dm^z$ is bounded away from 0 (Lemma C.9).

Fix $\epsilon > 0$, and pick $\delta = \frac{\epsilon}{8\|\Psi\|_\infty}$. For sample size N sufficiently large, we have that

$$P\left(V_{1,\delta} > \left(\sup_{\theta, q \in [\delta, 1-\delta]} \frac{4}{\int f_\varepsilon\left(\varepsilon\left(q, X_1; \theta\right)\right) f_\varepsilon\left(\varepsilon\left(q, X_2; \theta\right)\right) f_\eta\left(\eta\left(q, Z; \theta\right)\right) dm^x dm^x dm^z} \right)^{-1} \epsilon \right) < \epsilon$$

and $P\left(V_{2,\delta} > \frac{\epsilon}{4\|\Psi\|_\infty}\right) < \epsilon$. This implies that

$$P\left(\sup_{\theta} |\psi(\theta) - \psi_N(\theta; m_N^x, m_J^z)| > \epsilon\right) < \epsilon$$

proving the desired uniform convergence in probability. $\square$

C.2.4 Proof of Theorem 3.2

Fix $\epsilon > 0$, and choose δ such that $\|\psi(\theta) - \psi(\theta_0)\|_W < \frac{\delta}{2} \Rightarrow \|\theta - \theta_0\| < \epsilon$.

By Propositions 3.7 and 3.8 and the continuous mapping theorem, $\|\psi_N - \psi_N(\theta)\|_W$ converges in (outer) probability to $\|\psi - \psi(\theta)\|_W$ uniformly in θ.

Note that $\|\psi - \psi(\theta_0)\|_W = 0$.It follows that for sufficiently large N,

$$P\left(\sup_{\theta\in\Theta} |\|\psi_N - \psi_N(\theta)\|_W - \|\psi - \psi(\theta)\|_W| > \frac{\delta}{2}\right) < \epsilon,$$

so with probability at least $1 - \epsilon$,

$$\begin{aligned} \|\psi - \psi(\theta)\|_W &\leq \|\psi_N - \psi_N(\theta)\|_W + \frac{\delta}{2} \\ &\leq \delta \end{aligned}$$

However, by Assumption 3.6, this implies $\left\|\widehat{\theta}_N - \theta_0\right\| < \epsilon$. It follows that $\widehat{\theta}_N$ converges in probability to θ_0, proving consistency of the estimator.

C.3 Parametric Bootstrap

Let $\{z_j\}_{j=1}^J$ be a sample of firm characteristics and $\{x_i\}_{i=1}^N$ denote a sample of worker characteristics. The parametric bootstrap for the estimate $\hat{\theta} = \arg\min_{\theta\in\Theta} \hat{Q}_N(\theta)$ is constructed by the following procedure for $b = \{1, \ldots, 500\}$

1. Sample J firms with replacement from the empirical sample $\{z_j\}_{j=1}^J$. Denote this sample with $\left\{z_j^b\right\}_{j=1}^J$.

2. Draw N^b workers with replacement from the empirical sample $\{x_i\}_{i=1}^N$, where $N^b = \sum c_j^b$ and c_j^b is capacity of the j-th sampled firm in the bootstrap sample.

3. Simulate the unobservables ε_j^b and η_i^b.

4. Compute the quantities v_i^b and u_j^b at $\hat{\theta}$ from equations (3.12a) and (3.12b). For the model with preference heterogeneity, compute u_{ij}^b as in equation (3.14).

5. Compute a pairwise stable match for the bootstrap sample.

6. Compute $\hat{\theta}_b = \arg\min_{\theta \in \Theta} \hat{Q}_N^b(\theta)$ using the bootstrap pairwise stable match and an independent set of simulations for $\hat{Q}_N^b(\theta)$.

References

Abdulkadiroglu, A., Agarwal, N., and Pathak, P. A. (2012). Sorting and Welfare Consequences of Coordinated School Admissions: Evidence from New York City.

Abdulkadiroğlu, A., Angrist, J., and Pathak, P. (2011). The elite illusion: Achievement effects at boston and new york exam schools. NBER Working paper 17264.

Abdulkadiroğlu, A., Che, Y.-K., and Yasuda, Y. (2012). Expanding choice in school choice. Working paper, Duke University.

Abdulkadiroğlu, A., Pathak, P. A., and Roth, A. E. (2005). The new york city high school match. *American Economic Review, Papers and Proceedings*, 95:364–367.

Abdulkadiroğlu, A., Pathak, P. A., and Roth, A. E. (2009). Strategy-proofness versus efficiency in matching with indifferences: Redesigning the new york city high school match. *American Economic Review*, 99(5):1954–1978.

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

Abulkadiroglu, A., Hu, W., and Pathak, P. A. (2012). Small schools and student achievement: Evidence from new york city's reforms. Unpublished working paper, MIT.

Agarwal, N. (2012). An empirical model of the medical match. Unpublished working paper, Harvard.

Agarwal, N. and Diamond, W. F. (2013). Identification and Estimation in Two-sided Matching Markets.

Akkus, O., Cookson, A., and Hortacsu, A. (2012). The Determinants of Bank Mergers: A Revealed Preference Analysis.

Anderson, G. F. (1996). What Does Not Explain the Variation in the Direct Costs of Graduate Medical Education. *Academic Medicine*, 71(2):164–9.

Azevedo, E. M. and Leshno, J. D. (2011). Can we make school choice more efficient? an example. Unpublished mimeo, Harvard University.

Bailey, J. M. (2010). Health Care Reform, What's in it? Rural Communities and Medical Care. Technical report, Center for Rural Affairs.

Bajari, P. and Fox, J. T. (2005). Measuring the Efficiency of an FCC Spectrum Auction.

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

Becker, G. S. (1973). A Theory of Marriage: Part I. *Journal of Political Economy*, 81(4):813 – 46.

Becker, G. S. (1975). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education, 2nd ed*. National Bureau of Economic Research, Inc, New York.

Beggs, S., Cardell, S., and Hausman, J. (1981). Assessing the potential demand for electric cars. *Journal of Econometrics*, 17(1):1–19.

Ben-Porath, Y. (1967). The Production of Human Capital and the Life Cycle of Earnings. *Journal of Political Economy*.

Berry, S. (1994). Estimating discrete-choice models of product differentiation. *Rand Journal of Economics*, 25(2):242–262.

Berry, S. T., Levinsohn, J., and Pakes, A. (1995). Automobile Prices in Market Equilibrium. *Econometrica*, 63(4):841 – 890.

Berry, S. T. and Pakes, A. (2007). The Pure Characteristics Demand Model. *International Economic Review*, 48(4):1193–1225.

Blundell, R. and Powell, J. (2003). Endogeneity in Nonparametric and Semiparametric Regression Models. In Dewatripont, M., Hansen, L., and Turnovky, S., editors, *Advances in Economics and Econometrics*, chapter 8, pages 312–357.

Boyd, D., Lankford, H., Loeb, S., and Wyckoff, J. (2003). Analyzing the Determinants of the Matching Public School Teachers to Jobs: Estimating Compensating Differentials in Imperfect Labor Markets.

Breusch, T. S. and Pagan, A. R. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica*, 47(5):1287 – 94.

Brooks, R. G., Walsh, M., Mardon, R. E., Lewis, M., and Clawson, A. (2002). The Roles of Nature and Nurture in the Recruitment and Retention of Primary Care Physicians in Rural Areas: A Review of the Literature. *Academic Medicine*, 77(8):790–8.

Bulow, J. and Levin, J. (2006). Matching and Price Competition. *The American Economic Review*, 96(3):652 – 668.

Camina, E. (2006). A generalized assignment game. *Mathematical Social Sciences*, 52(2):152 – 161.

Chiappori, P.-A. and Komunjer, I. (2008). Correct Specification and Identification of Nonparametric Transformation Models. Technical report.

Chiappori, P.-A., Salanié, B., and Weiss, Y. (2011). Partner Choice and the Marital College Premium.

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

Christakis, N. A., Fowler, J. H., Imbens, G. W., and Kalyanaraman, K. (2010). An Empirical Model for Strategic Network Formation.

Clark, S. (2006). The Uniqueness of Stable Matchings. *The B.E. Journal of Theoretical Economics*, contr. 6(1):8.

Clotfelter, C. (1999). Public school segregation in metropolitan areas. *Land Economics*, 75(4):487–504.

Conlon, C. T. and Mortimer, J. H. (2010). Demand Estimation Under Incomplete Product Availability.

Crawford, V. P. (2008). The Flexible-Salary Match: A proposal to increase the salary flexibility of the National Resident Matching Program. *Journal of Economic Behavior & Organization*, 66(2):149–160.

Dubins, L. E. and Freedman, D. A. (1981). Machiavelli and the gale-shapley algorithm. *American Mathematical Monthly*, 88:485–494.

Engberg, J., Epple, D., Imbrogno, J., Sieg, H., and Zimmer, R. (2010). Bounding the treatment effects of education programs that have lotteried admission and selective attrition. Unpublished working paper, University of Pennsylvania.

Epple, D. and Romano, R. (1998). Competition between private and public schools, vouchers, and peer group effects. *American Economic Review*, 88(1):33–62.

Erdil, A. and Ergin, H. (2008). What's the matter with tie-breaking? improving efficiency in school choice. *American Economic Review*, 98:669–689.

Fox, J. T. (2008). Estimating Matching Games with Transfers.

Fox, J. T. (2009). Structural Empirical Work Using Matching Models. In Durlauf, S. N. and Blume, L. E., editors, *New Palgrave Dictionary of Economics*. Online edition.

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

Galichon, A. and Salanie, B. (2010). Matching with Trade-offs: Revealed Preferences over Competing Characteristics.

Gendar, A. (2000). Simpler hs entry eyed ed board aims to trim selection process, improve schools. *New York Daily News, October 21*.

Gentile Jr., T. C. and Buckley, D. (2009). Medicare Reimbursement and Graduate Medical Education. In Levine, J., editor, *Guide to Medical Education in the Teaching Hospital*. Association for Hospital Medical Education, 4th edition.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, first edition.

Gordon, N. and Knight, B. (2009). A spatial merger estimator with an application to school district consolidation. *Journal of Public Economics*, pages 752 – 765.

Gourieroux, C. and Monfort, A. (1997). *Simulation-based Econometric Methods*. Oxford University Press, New York.

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

Hastings, J., Kane, T. J., and Staiger, D. O. (2005). Parental preferences and school competition: Evidence from a public school choice plan. NBER Working Paper, 11805.

Hastings, J., Kane, T. J., and Staiger, D. O. (2009). Heterogenous preferences and the efficacy of public school choice. Working paper, Yale University.

Hastings, J. and Weinstein, J. M. (2008). Information, school choice and academic achievement: Evidence from two experiments. *Quarterly Journal of Economics*, 123(4):1373–1414.

He, Y. (2012). Gaming the boston mechanism in beijing. Unpublished mimeo, Toulouse.

Heckman, J. J., Lochner, L. J., and Todd, P. E. (2003). Fifty Years of Mincer Earnings Regressions.

Heckman, J. J. and Robb, R. J. (1985). Alternative methods for evaluating the impact of interventions: An overview. *Journal of Econometrics*, 30(1-2):239–267.

Hemphill, C. (2007). *New York City's Best Public High Schools: A Parent's Guide, Third Edition*. Teachers College Press.

Hemphill, C. and Nauer, K. (2009). The new marketplace: How small-school reforms and school choice have reshaped new york city's high schools. Center for New York City Affairs, Milano the New School for Management and Urban Policy.

Herszenhorn, D. M. (2004). Council members see flaws in school-admissions plan. New York Times, Metropolitan Desk, November 19.

Hoxby, C. (2003). School choice and school productivity (or, could school choice be a rising tide that lifts all boats). In Hoxby, C., editor, *The Economics of School Choice*. Chicago: University of Chicago Press.

Imbens, G. W. (2007). Nonadditive Models with Endogenous Regressors. In Blundell, R., Newey, W., and Persson, T., editors, *Advances in Economics and Econometrics, Vol III*, chapter 2, pages 17 – 46.

Imbens, G. W. and Newey, W. K. (2009). Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity. *Econometrica*, 77(5):1481 – 1512.

Jacob, B. and Lefgren, L. (2007). What do parents value in education? an empirical investigation of parents' revealed preferences for teachers. *Quarterly Journal of Economics*, 122(4):1603–1637.

Johnson, S. G. (2011). The NLopt nonlinear-optimization package.

Jung et.al. v AAMC et.al. (2002). Class Action Complaint, No. 02-CV-00873, D.D.C. May 5, 2002.

Kagel, J. and Roth, A. E. (2000). The dynamics of reorganization in matching markets: A laboratory experiment motivated by a natural experiment. *Quarterly Journal of Economics*, pages 201–235.

Kamada, Y. and Kojima, F. (2010). Improving Efficiency in Matching Markets with Regional Caps: The Case of the Japan Residency Matching Program.

Kelso, A. S. J. and Crawford, V. P. (1982). Job Matching, Coalition Formation, and Gross Substitutes. *Econometrica*, 50(6):1483 – 1504.

Kerr, P. (2003). Parents perplexed by high school admissions rule: Reply. New York Times, November 3.

Kesten, O. (2010). School choice with consent. *Quarterly Journal of Economics*, 125(3):1297–1348.

Kesten, O. and Kurino, M. (2012). On the (im)possibility of improving upon the student-proposing deferred acceptance mechanism. WZB Discussion Paper SP II 2012-202.

Kling, J. R., Mullainathan, S., Shafir, E., Vermuelen, L. C., and Wrobel, M. V. (2012). Comparison friction: Experimental evidence from medicare drug plans. *Quarterly Journal of Economics*, 127:199–235.

Kojima, F. (2007). Matching and Price Competition: Comment. *American Economic Review*, 97(3):1027 – 1031.

Kojima, F., Pathak, P., and Roth, A. E. (2010). Matching with Couples: Stability and Incentives in Large Markets.

Kojima, F. and Pathak, P. A. (2009). Incentives and Stability in Large Two-Sided Matching Markets. *American Economic Review*, 99(3):608 – 27.

Kolaczyk, E. D. (2009). *Statistical Analysis of Network Data: Methods and Models*. Springer.

Logan, J. A., Hoff, P. D., and Newton, M. A. (2008). Two-Sided Estimation of Mate Preferences for Similarities in Age, Education, and Religion. *Journal of the American Statistical Association*, pages 559 – 569.

McFadden, D. (1989). A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration. *Econometrica*, 57(5):995 – 1026.

Meier, D. (2002). *The Power of their Ideas: Lessons for America from a Small School in Harlem*. Beacon Press.

Menzel, K. (2011). Robust Decisions For Incomplete Structural Models Of Social Interactions.

Mincer, J. (1974). *Schooling, Experience, and Earnings*. National Bureau of Economic Research, Inc, New York.

Mortensen, D. T. and Pissarides, C. A. (1994). Job Creation and Job Destruction in the Theory of Unemployment. *Review of Economic Studies*, 61(3):397 – 415.

Newhouse, J. P. and Wilensky, G. R. (2001). Paying For Graduate Medical Education: The Debate Goes On. *Health Affairs*, 20(2):136–147.

Niederle, M. and Roth, A. E. (2003a). Relationship Between Wages and Presence of a Match in Medical Fellowships. *JAMA, Journal of the American Medical Association*, 290(9).

Niederle, M. and Roth, A. E. (2003b). Unraveling reduced mobility in a labor market: Gastroenterology with and without a centralized match. *Journal of Political Economy*, 111:1342–1352.

Niederle, M. and Roth, A. E. (2009). The Effects of a Centralized Clearinghouse on Job Placement, Wages, and Hiring Practices. In David Autor, editor, *NBER Chapters*, pages 235 – 271.

Niederle, M. and Yariv, L. (2009). Decentralized Matching with Aligned Preferences.

Pakes, A. and Pollard, D. (1989). Simulation and the Asymptotics of Optimization Estimators. *Econometrica*, 57(5):1027 – 57.

Pathak, P. A. and Sönmez, T. (2008). Leveling the playing field: Sincere and sophisticated players in the boston mechanism. *American Economic Review*, 98(4):1636–1652.

Pathak, P. A. and Sönmez, T. (2011). School admissions reform in chicago and england: Comparing mechanisms by their vulnerability to manipulation. forthcoming, *American Economic Review*.

Pennell, H., West, A., and Hind, A. (2006). Secondary school admissions in london. Centre for Education Research, LSE CLARE Working paper No. 19.

Petrin, A. K. and Train, K. E. (2010). A control function approach to endogeneity in consumer choice models. *Journal of Marketing Research*, 47(1):3–13.

Postel-Vinay, F. and Robin, J.-M. (2002). Equilibrium Wage Dispersion with Worker and Employer Heterogeneity. *Econometrica*, 70(6):2295 – 2350.

Rabinowitz, H. K., Diamond, J. J., Markham, F. W., and Wortman, J. R. (2008). Medical school programs to increase the rural physician supply: a systematic review and projected impact of widespread replication. *Academic Medicine*, 83(3):235–43.

Ravitch, D. (2011). *The Death and Life of the Great American School System*. Basic Books.

Rosen, S. (1972). Learning and experience in the labor market. *Journal of Human Resources*, 7(3):326–342.

Rosen, S. (1987). The theory of equalizing differences. In Ashenfelter, O. and Layard, R., editors, *Handbook of Labor Economics*, volume 1 of *Handbook of Labor Economics*, chapter 12, pages 641–692. Elsevier.

Rosenblatt, R. A., Hagopian, A., Andrilla, C. H. A., and Hart, G. (2006). Will Rural Family Medicine Residency Training Survive? *Family medicine*, 38(10):706–11.

Rosenblatt, R. A. and Hart, L. G. (2000). Physicians and Rural America. *The Western journal of medicine*, 173(5):348–51.

Roth, A. E. (1982). The economics of matching: Stability and incentives. *Mathematics of Operations Research*, 7:617–628.

Roth, A. E. (1984). The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory. *Journal of Political Economy*, 92(6):991–1016.

Roth, A. E. and Peranson, E. (1999). The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design. *American Economic Review*, 89(4):748–780.

Roth, A. E. and Sotomayor, M. A. O. (1992). *Two-Sided Matching*. Cambridge University Press.

Roth, A. E. and Vande Vate, J. H. (1990). Random Paths to Stability in Two-Sided Matching. *Econometrica*, 58(6):1475 – 80.

Roth, A. E. and Xing, X. (1994). Jumping the Gun: Imperfections and Institutions Related to the Timing of Market Transactions. *American Economic Review*, 84(4):992–1044.

Roth, A. E. and Xing, X. (1997). Turnaround time and bottlenecks in market clearing: Decentralized matching in the market for clinical psychologists. *Journal of Political Economy*, 105:284–329.

Rothstein, J. (2006). Good principals or good peers: Parental valuation of school characteristics, tiebout equilibrium, and the incentive effects of competition among jurisdictions. *American Economic Review*, 96(4):1333–1350.

Rowan, T. H. (1990). *Functional Stability Analysis Of Numerical Algorithms*. PhD thesis.

Schneider, M., Teske, P., and Marschall, M. (2002). *Choosing Schools: Consumer Choice and the Quality of American Schools*. Princeton University Press.

Shapley, L. and Scarf, H. (1974). On cores and indivisibility. *Journal of Mathematical Economics*, 1:23–28.

Shapley, L. and Shubik, M. (1971). The assignment game i: The core. *International Journal of Game Theory*, 1(1):111–130.

Shimer, R. and Smith, L. (2000). Assortative Matching and Search. *Econometrica*, 68(2):343 – 370.

Signer, M. M. (2012). 2012 Main Residency Match and SOAP. Technical report, National Resident Matching Program.

Sorensen, M. (2007). How Smart Is Smart Money? A Two-Sided Matching Model of Venture Capital. *Journal of Finance*, 62(6):2725 – 2762.

Sotomayor, M. (1999). The Lattice Structure of the Set of Stable Outcomes of the Multiple Partners Assignment Game. *International Journal of Game Theory*, 28(4):567 – 583.

Stern, S. (2004). Do Scientists Pay to Be Scientists? *Management Science*, 50(6):835–853.

Stock, J. H., Wright, J. H., and Yogo, M. (2002). A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments. *Journal of Business & Economic Statistics*, 20(4):518–29.

Talley, R. C. (1990). Graduate medical education and rural health care. *Academic medicine*, 65(12 Suppl):S22–5.

Teulings, C. N. (1995). The Wage Distribution in a Model of the Assignment of Skills to Jobs. *Journal of Political Economy*, 103(2):280 – 315.

Thomas Lemieux (2006). The Mincer Equation Thirty Years After Schooling, Experience, and Earnings. In *Jacob Mincer: A Pioneer of Modern Labor Economics*.

Uetake, K. and Watanabe, Y. (2012). Entry by Merger: Estimates from a Two-Sided Matching Model with Externalities.

Urquiola, M. (2005). Does school choice lead to sorting? evidence from tiebout variation. *American Economic Review*, 95(4):1310–1326.

van der Vaart, A. and Wellner, J. (2000). *Weak Convergence and Empirical Processes: With Applications to Statistics (Springer Series in Statistics)*. Springer, corrected edition.

Vanacore, A. (2012). Centralized enrollment in recovery school district gets first tryout. *Times-Picayune, New Orleans, April 16*.

Weese, E. (2008). Political Mergers as Coalition Formation: Evidence from Japanese Municipal Amalgamations.

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*, 48(4):817 – 38.

Williams, J. (2004). Hs delay pinches parents. *New York Daily News, March 25*.

Yang, S., Chen, Y., and Allenby, G. M. (2003). Bayesian Analysis of Simultaneous Demand and Supply. *Quantitative Marketing and Economics*, 1(3):251–275.