NBER WORKING PAPER SERIES

AN EMPIRICAL MODEL OF THE MEDICAL MATCH

Nikhil Agarwal

Working Paper 20767
http://www.nber.org/papers/w20767

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2014

I am grateful to my advisors Ariel Pakes, Parag Pathak, Susan Athey and Al Roth for their constant support and guidance. I thank Atila Abdulkadiroglu, Raj Chetty, David Cutler, Rebecca Diamond, William Diamond, Adam Guren, Guido Imbens, Dr. Joel Katz, Larry Katz, Greg Lewis, Jacob Leshno, Julie Mortimer, Joseph Newhouse, Mark Shepard, Dr. Debra Weinstein and workshop participants at several universities for helpful discussions, suggestions and comments. Data acquisition for this project was funded by the Lab for Economic Applications and Policy and the Kuznets Award. Financial support from the NBER Non-profit Fellowship and Yahoo! Key Scientific Challenges Program is gratefully acknowledged. Computations for this paper were run on the Odyssey cluster supported by the FAS Science Division Research Computing Group at Harvard University. Email: agarwaln@mit.edu. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

An Empirical Model of the Medical Match
Nikhil Agarwal
NBER Working Paper No. 20767
December 2014
JEL No. C51,C78,D47,J41,J44,L44

ABSTRACT

This paper develops a framework for estimating preferences in two-sided matching markets with non-transferable utility using only data on observed matches. Unlike single-agent choices, matches depend on the preferences of other agents in the market. I use pairwise stability together with a vertical preference restriction on one side of the market to identify preference parameters for both sides of the market. Recovering the distribution of preferences is only possible in an environment with many-to-one matching. These methods allow me to investigate two issues concerning the centralized market for medical residents. First, I examine the antitrust 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 product 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 and supply policies aimed at increasing the number of residents training in rural areas while accounting for general equilibrium effects from the matching market. I find that financial incentives increase the quality, but not the number of rural residents. Quantity regulations increase the number of rural trainees, but the impact on resident quality depends on the design of the intervention.

Nikhil Agarwal
MIT
Department of Economics, E17-224
77 Massachusetts Ave
Cambridge, MA 02139
and NBER
agarwaln@mit.edu

A data appendix is available at:
http://www.nber.org/data-appendix/w20767

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 residency 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

[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.

the presence of a centralized match (Niederle and Roth, 2003, 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

[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.

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 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

[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.

medical school prestige decreases with the correlation of human capital with medical school prestige. Note that it is only possible to calculate the 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 Heckman and Robb, 1985; Blundell and Powell, 2003; 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 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

[4] Agarwal and Diamond (in progress) 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.

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 (2003, 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, Salanié, and Weiss (2010), Galichon and Salanie (2012), among others, build on the seminal work of Choo and Siow (2006) for studying transferable utility models of the marriage market in which an aggregate surplus is split between spouses. Fox (2010) proposes a different approach for estimation, also for the transferable utility case, with applications in Bajari and Fox (2013), 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 (Weese, 2014; Gordon and Knight, 2009; Akkus, Cookson, and Hortacsu, 2012; Uetake and Watanabe, 201). 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, Hoff, and Newton (2008) and Boyd et al. (2013), although in decentralized markets, with the goal of measuring preferences for various characteristics. Logan, Hoff, and Newton (2008) proposes a Bayesian method for estimating preferences for mates in a marriage market with no monetary transfers. Boyd et al. (2013) uses the method of simulated

[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; Roth and Xing, 1994; Shimer and Smith, 2000; Postel-Vinay and Robin, 2002).

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 (2014) 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 (Rosen, 1972; Becker, 1975). 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). Sections 3 through 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 8 and 9 respectively. All technical details are relegated to appendices.

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

[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 E. Data on matches from the Graduate Medical

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 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 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 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.

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 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.

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 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]

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

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

structure to infer preferences. I defer the discussion of the many-to-one aspect to Section 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 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 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 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 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]

To highlight the geographical sorting observed in the data, Table 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

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

model of the market that is estimated using these patterns in the data.

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 5.

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 $\mu_t : \mathcal{N}_t \rightarrow \mathcal{J}_t$. 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 $i' \in \mu_t^{-1}(j)$, $\mu_t^{-1}(j) \succeq_j (\mu_t(j) \setminus \{i'\}) \cup \{i\}$
   - If $|\mu(j)| < c_j$, then $\mu_t^{-1}(j) \succeq_j \mu_t^{-1}(j) \cup \{i\}$.

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 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.

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 $U\left(z_{jt}, \eta_{jt}, w_{jt}, \beta_i; \theta\right)$ 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

[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).

[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, Pathak, and Roth (2013) suggest that stable matches exist in large markets if the fraction of couples is small.

characteristics demand model of Berry and Pakes (2007) for this indirect utility:

$$u_{ijt} = z_{jt}\beta_i^z + w_{jt}\beta_i^w + \xi_{jt}. \tag{1}$$

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:

$$\beta_i = x_i\Pi + \eta_i. \tag{2}$$

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.

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

[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 impor-

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 "pecking 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

$$h_i = x_i\alpha + \varepsilon_i, \tag{3}$$

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 6.3.[17] Finally, Section 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

tant 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.

[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 D.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.

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 (2014) 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 5, and does not require a reconsideration of arguments presented here.

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 $E\left[u_j|z_j=0\right]=0$ and $E\left[h_i|x_i=0\right]=0$.

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 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 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 (2014) 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.

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

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]

[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

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 (2014) 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.

4.2.1 Descriptive Statistics from Many-to-One Matching

Table 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 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 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 5 might have suggested.

program are very similarly qualified, the observable quality of two doctors at the same program act like noisy measures of their identical true quality.

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 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 (2013) 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.

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]

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 Heckman and Robb, 1985; Blundell and Powell, 2003; 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 :

$$w_{jt} = z_{jt}\gamma + r_{jt}\tau + \nu_{jt}, \tag{4}$$

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 (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 (4) using OLS, treat it as any other observed characteristic. As noted earlier, we need to allow for correlation between ν_{jt} and

[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, Levinsohn, and Pakes (1995), where an inversion can be used to to estimate a variable with a separable form in the unobserved characteristic and the endogenous variable.

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

$$\xi_{jt} = \kappa\nu_{jt} + \sigma\zeta_{jt}, \tag{5}$$

where $\zeta_{jt} \sim N(0,1)$ is drawn independently of ν_{jt} and (κ, σ) are unknown parameters. Substitute equation (5) to re-write equation (1) as

$$u_{ijt} = z_{jt}\beta_i^z + w_{jt}\beta_i^w + \kappa\nu_{jt} + \sigma\zeta_{jt}. \tag{6}$$

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, Chen, and Allenby (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 (5) and (6) in settings with greater variation in the endogenous variable.

5.2 Instrument

Table 7 presents regression estimates of equation (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 7, 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

[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).

$$r_j = \frac{\sum_{k \in G_j} fte_k \times rr_k}{\sum_{k \in G_j} fte_k}, \tag{7}$$

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 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 7 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

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

[24] Figure G.2 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 G.3 presents a scatter plot of the salary distribution against fitted values. The plot shows little evidence of heteroskedasticity.

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

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.

6 Estimation

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

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 (4) using OLS to construct the residual

$$\hat{\nu}_{jt} = w_{jt} - z_{jt}\hat{\gamma} - r_{jt}\hat{\tau}. \tag{8}$$

Replacing this estimate in equation (6), we get

$$u_{ijt} \approx z_{ijt}\beta_i^z + w_{jt}\beta_i^w + \kappa\hat{\nu}_{jt} + \sigma\zeta_{jt}, \tag{9}$$

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 (6), (2) and (3). 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

$$\left\|\hat{m} - \hat{m}^S(\theta)\right\|_W^2 = \left(\hat{m} - \hat{m}^S(\theta)\right)' W \left(\hat{m} - \hat{m}^S(\theta)\right), \tag{10}$$

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

can be rejected.

W is a matrix of weights described in Section 6.4. Additional details on the estimator and the optimization algorithm are in Appendix A.[26]

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

$$\hat{m}_{t,ov} = \frac{1}{N_t} \sum_{i \in \mathcal{N}_t} 1\{\mu_t(i) = j\} x_i z_{jt}. \tag{11}$$

2. The within-program variance of resident observables. For each scalar $x_{1,i}$:

$$\hat{m}_{t,w} = \frac{1}{N_t} \sum_{i \in \mathcal{N}_t} \left(x_{1,i} - \frac{1}{\left|\mu_t^{-1}(\mu_t(i))\right|} \sum_{i' \in \mu_t^{-1}(\mu_t(i))} x_{1,i'} \right)^2. \tag{12}$$

3. The covariance between resident characteristics and the average characteristics of a resident's peers. For every pair of scalars $x_{1,i}$ and $x_{2,i}$:

$$\hat{m}_{t,p} = \frac{1}{N_t} \sum_{i \in \mathcal{N}_t} x_{1,i} \frac{1}{\left|\mu_t^{-1}(\mu_t(i))\right| - 1} \sum_{i' \in \mu_t^{-1}(\mu_t(i)) \setminus \{i\}} x_{2,i'}. \tag{13}$$

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 3 and 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 na-

[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 A.

ture of the market.[27] Section 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]

6.3 Simulating a Match

Under the parametric assumptions made on ζ_{jt}, ε_i, and η_i in Section 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 $\left\{\zeta_{js}\right\}_{j=1}^{J}$ indexed by s, calculate

$$h_{is} = x_i\alpha + \varepsilon_{is}, \tag{14}$$

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 (11) to (13). The optimization routine keeps a fixed set of simulation draws of unobservable characteristics for computing moments at different values of θ.

[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 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.

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 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]

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 (2014) 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 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

[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 $O\left(n^3\right)$.

programs with characteristics (x_i, ε_i) and $\left(z_{jt}, \xi_{jt}\right)$ 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 Kolaczyk, 2009; Christakis et al., 2010). 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):

$$\hat{\Sigma} = \left(\hat{\Gamma}' W \hat{\Gamma}\right)^{-1} \hat{\Gamma}' W \left(\hat{V} + \frac{1}{S} \hat{V}^S\right) W' \hat{\Gamma} \left(\hat{\Gamma}' W \hat{\Gamma}\right)^{-1}, \tag{15}$$

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 A 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 $\left(\varepsilon_{i,b},\ \nu_{i,b},\ \xi_{jt,b}\right)$ 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 A.

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 $\hat{V}^{-1}$.

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 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 briefly discusses the underlying parameters, which are not economically intuitive, and robustness using estimates from additional models.

7.1 Preference Estimates

Panel A.1 of Table 8 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, 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 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.

7.1.1 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 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 7 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 bias in the salary coefficient using specification (3).

7.1.2 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 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 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

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

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.

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 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

[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.

not leading to poor predictions of the sorting patterns in the market. Therefore, I am comfortable using estimates as basis of counterfactual analysis.

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 (2003, 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 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

[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.

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.

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

$$f(h_k, q_k) - w_k \geq f(h_i, q_k) - w_i + a(q_k - q_i). \tag{16}$$

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 $w_i - a(q_k - q_i)$.

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 C.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

[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).

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

$$w_k = -aq_k + \sum_{i=2}^{k} \left[f(h_i, q_i) - f(h_{i-1}, q_i) \right]. \tag{17}$$

Resident 1 receives her product of labor $f(h_1, q_1)$ (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, $f(h_2, q_2) - f(h_1, q_2)$, 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 (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

[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 (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.

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 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 5 stated and proved in Appendix C.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 $f(h_k, q_k)$.

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

$$w_k = \bar{f}(h_k) - aq_k. \tag{18}$$

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

$$w_k = -aq_k. \tag{19}$$

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 1 shows that in all cases resident k pays the implicit tuition aq_k. Equilibrium wages given in equations (18) and (19) highlight that the side of the market that owns the factor determining

[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).

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.

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 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 C.4.[39] With a slight abuse of notation, let the total surplus from the pair $(i,\ j)$ be $a^f_{ij} = u_{ij} + f(h_i) \geq 0$.[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 $a^{\tilde{f}}_{ij} = u_{ij} + \tilde{f}(h_i) \geq 0$ in terms

[39] In the general formulation, I assume that the total output from a team of residents $\left(h_1, \ldots, h_{q_j}\right)$ is $F\left(h_1, \ldots, h_{q_j}\right) = \sum_{k=1}^{q_j} f(h_k)$, 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.

of the equilibria of the game with surplus a_{ij}^{f}.

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

Proof. See Appendix C.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 C.3 describes the procedure in more detail. All calculations are done with the 2010-2011 sample of the data.

[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.

8.3 Estimates of Implicit Tuition

Estimates presented in Section 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 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 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 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

[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.

[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 F.

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 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.

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 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 intrinsically 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 (2003, 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.

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 D.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 D.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

[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 (Rosenblatt and Hart, 2000; Rabinowitz et al., 2008).

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).

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 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.

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 9).

[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)."

Panel A of Table 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 to the quality of doctors in urban areas. This is consistent with the reduced form evidence presented in Table 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 analysis

[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.

based on unilateral salary increases substantially deviates from this prediction. For smaller interventions we expect general equilibrium effects to be less pronounced. In Appendix D.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 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.

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 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.

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.

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 Hastings, Kane, and Staiger, 2009; Abdulkadiroglu, Agarwal, and Pathak, 2014, 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 (2003, 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.

References

Abdulkadiroglu, A., N. Agarwal, and P. A. Pathak (2014). The Welfare Effects of Coordinated School Assignment: Evidence from the NYC High School Match.

Agarwal, N. and W. F. Diamond (2014). Identification and Estimation in Two-sided Matching Markets. *Cowles Foundation Discussion Paper No. 1905*.

Agarwal, N. and W. F. Diamond (in progress). Identification and Estimation in Two-sided Matching Markets.

Akkus, O., A. Cookson, and A. Hortacsu (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.

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 J. T. Fox (2013). Measuring the Efficiency of an FCC Spectrum Auction. *American Economic Journal: Microeconomics 5*(1), 100–146.

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.* New York: National Bureau of Economic Research, Inc.

Ben-Porath, Y. (1967). The Production of Human Capital and the Life Cycle of Earnings. *Journal of Political Economy*.

Berry, S. T. (1994). Estimating Discrete-Choice Models of Product Differentiation. *RAND Journal of Economics 25*(2), 242 – 262.

Berry, S. T., J. Levinsohn, and A. Pakes (1995). Automobile Prices in Market Equilibrium. *Econometrica 63*(4), 841 – 890.

Berry, S. T. and A. Pakes (2007). The Pure Characteristics Demand Model. *International Economic Review 48*(4), 1193–1225.

Blundell, R. and J. Powell (2003). Endogeneity in Nonparametric and Semiparametric Regression Models. In M. Dewatripont, L. Hansen, and S. Turnovky (Eds.), *Advances in Economics and Econometrics*, Chapter 8, pp. 312–357.

Boyd, D., H. Lankford, S. Loeb, and J. Wyckoff (2013). Analyzing the Determinants of the Matching Public School Teachers to Jobs: Estimating Compensating Differentials in Imperfect Labor Markets. *Journal of Labor Economics 31*(1), 83–117.

Breusch, T. S. and A. R. Pagan (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica 47*(5), 1287 – 94.

Brooks, R. G., M. Walsh, R. E. Mardon, M. Lewis, and A. Clawson (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 J. Levin (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., B. Salanié, and Y. Weiss (2010). Partner Choice and the Marital College Premium. *Columbia University, mimeo*.

Choo, E. and A. Siow (2006). Who Marries Whom and Why. *Journal of Political Economy 114*(1), 175–201.

Christakis, N. A., J. H. Fowler, G. W. Imbens, and K. Kalyanaraman (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.

Conlon, C. T. and J. H. Mortimer (2013). Demand Estimation Under Incomplete Product Availability. *American Economic Journal: Microeconomics 5*(4), 1–30.

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.

Fox, J. T. (2009). Structural Empirical Work Using Matching Models. In S. N. Durlauf and L. E. Blume (Eds.), *New Palgrave Dictionary of Economics* (Online ed.).

Fox, J. T. (2010). Estimating Matching Games with Transfers. *University of Michigan, mimeo*.

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

Galichon, A. and B. Salanie (2012). CupidâĂŹs Invisible Hand: Social Surplus and Identification in Matching Models. *Columbia University, mimeo*.

Gentile Jr., T. C. and D. Buckley (2009). Medicare Reimbursement and Graduate Medical Education. In J. Levine (Ed.), *Guide to Medical Education in the Teaching Hospital* (4th ed.). Association for Hospital Medical Education.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning* (First ed.). Boston, MA: Addison-Wesley Longman Publishing Co., Inc.

Gordon, N. and B. Knight (2009). A spatial merger estimator with an application to school district consolidation. *Journal of Public Economics*, 752 – 765.

Gourieroux, C. and A. Monfort (1997). *Simulation-based Econometric Methods.* New York: Oxford University Press.

Hastings, J. S., T. J. Kane, and D. O. Staiger (2009). Heterogeneous Preferences and the Efficacy of Public School Choice.

Heckman, J. J., L. J. Lochner, and P. E. Todd (2003). Fifty Years of Mincer Earnings Regressions.

Heckman, J. J. and R. J. Robb (1985). Alternative methods for evaluating the impact of interventions: An overview. *Journal of Econometrics 30*(1-2), 239–267.

Imbens, G. W. (2007). Nonadditive Models with Endogenous Regressors. In R. Blundell, W. Newey, and T. Persson (Eds.), *Advances in Economics and Econometrics, Vol III*, Chapter 2, pp. 17 – 46.

Imbens, G. W. and W. K. Newey (2009). Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity. *Econometrica 77*(5), 1481 – 1512.

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.

Kamada, Y. and F. Kojima (2010). Improving Efficiency in Matching Markets with Regional Caps: The Case of the Japan Residency Matching Program.

Kelso, A. S. J. and V. P. Crawford (1982). Job Matching, Coalition Formation, and Gross Substitutes. *Econometrica 50*(6), 1483 – 1504.

Kojima, F. (2007). Matching and Price Competition: Comment. *American Economic Review 97*(3), 1027 – 1031.

Kojima, F., P. Pathak, and A. E. Roth (2013). Matching with Couples: Stability and Incentives in Large Markets. *Quarterly Journal of Economics 128*(4), 1585–1632.

Kojima, F. and P. A. Pathak (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., P. D. Hoff, and M. A. Newton (2008). Two-Sided Estimation of Mate Preferences for Similarities in Age, Education, and Religion. *Journal of the American Statistical Association*, 559 – 569.

McFadden, D. (1989). A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration. *Econometrica 57*(5), 995 – 1026.

Mincer, J. (1974). *Schooling, Experience, and Earnings.* New York: National Bureau of Economic Research, Inc.

Mortensen, D. T. and C. A. Pissarides (1994). Job Creation and Job Destruction in the Theory of Unemployment. *Review of Economic Studies 61*(3), 397 – 415.

Newhouse, J. P. and G. R. Wilensky (2001). Paying For Graduate Medical Education: The Debate Goes On. *Health Affairs 20*(2), 136–147.

Niederle, M. and A. E. Roth (2003). Relationship Between Wages and Presence of a Match in Medical Fellowships. *JAMA, Journal of the American Medical Association 290*(9).

Niederle, M. and A. E. Roth (2009). The Effects of a Centralized Clearinghouse on Job Placement, Wages, and Hiring Practices. In David Autor (Ed.), *NBER Chapters*, pp. 235 – 271.

Niederle, M. and L. Yariv (2009). Decentralized Matching with Aligned Preferences. *Stanford University, mimeo*.

Pakes, A. and D. Pollard (1989). Simulation and the Asymptotics of Optimization Estimators. *Econometrica 57*(5), 1027 – 57.

Petrin, A. K. and K. E. Train (2010). A control function approach to endogeneity in consumer choice models. *Journal of Marketing Research 47*(1), 3–13.

Postel-Vinay, F. and J.-M. Robin (2002). Equilibrium Wage Dispersion with Worker and Employer Heterogeneity. *Econometrica 70*(6), 2295 – 2350.

Rabinowitz, H. K., J. J. Diamond, F. W. Markham, and J. R. Wortman (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.

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 O. Ashenfelter and R. Layard (Eds.), *Handbook of Labor Economics*, Volume 1 of *Handbook of Labor Economics*, Chapter 12, pp. 641–692. Elsevier.

Rosenblatt, R. A., A. Hagopian, C. H. A. Andrilla, and G. Hart (2006). Will Rural Family Medicine Residency Training Survive? *Family medicine 38*(10), 706–11.

Rosenblatt, R. A. and L. G. Hart (2000). Physicians and Rural America. *The Western journal of medicine 173*(5), 348–51.

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 E. Peranson (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 M. A. O. Sotomayor (1992). *Two-Sided Matching.* Cambridge University Press.

Roth, A. E. and J. H. Vande Vate (1990). Random Paths to Stability in Two-Sided Matching. *Econometrica 58*(6), 1475 – 80.

Roth, A. E. and X. Xing (1994). Jumping the Gun: Imperfections and Institutions Related to the Timing of Market Transactions. *American Economic Review 84*(4), 992–1044.

Rowan, T. H. (1990). *Functional Stability Analysis Of Numerical Algorithms.* Ph. D. thesis.

Shapley, L. S. and M. Shubik (1971). The assignment game I: The core. *International Journal of Game Theory 1*(1), 111–130.

Shimer, R. and L. Smith (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., J. H. Wright, and M. Yogo (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 Y. Watanabe (201). Entry by Merger: Estimates from a Two-Sided Matching Model with Externalities. *Yale University, mimeo*.

Weese, E. (2014). Political Mergers as Coalition Formation: An Analysis of the Heisei Municipal Amalgamations. *Quantitative Economics, forthcoming*.

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica 48*(4), 817 – 38.

Yang, S., Y. Chen, and G. M. Allenby (2003). Bayesian Analysis of Simultaneous Demand and Supply. *Quantitative Marketing and Economics 1*(3), 251–275.

Table 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: 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.

Table 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 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.

Table 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.

Table 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.

Table 7: 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: Regression of a program's first year salary on program characteristics. 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). Columns (2-7) include dummy variables for programs with no NIH funding at major affiliates, for no NIH funding at minor affiliates, and a dummy for missing Medicare ID for the primary institution. All columns include a constant term. The Competitor Salary (Lagged) is the average of lagged salaries of other family practice residency programs in the geographic area of the program hospital. Therefore, columns (4) and (6) exclude data from the first year of the sample, 2003-2004. 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 (5-7), 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. Sample in all columns restricted to programs for which salary was not imputed from the regressions described in the data appendix. Significance at 90% (*), 95% (**) and 99% (***) confidence.

Table 8: 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 8: 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.

Table 9: Estimated Utility Distribution in First-Year Salary Equivalent

		Full Heterogeneity (1)		Geographic Heterogeneity (2)		Geo. Het. w/ Wage Instrument (3)	
	N	Stat	(s.e.)	Stat	(s.e.)	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 6.4. Statistics use the 2010-2011 sample.

Table 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 6.4. Standard errors in parenthesis.

Table 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

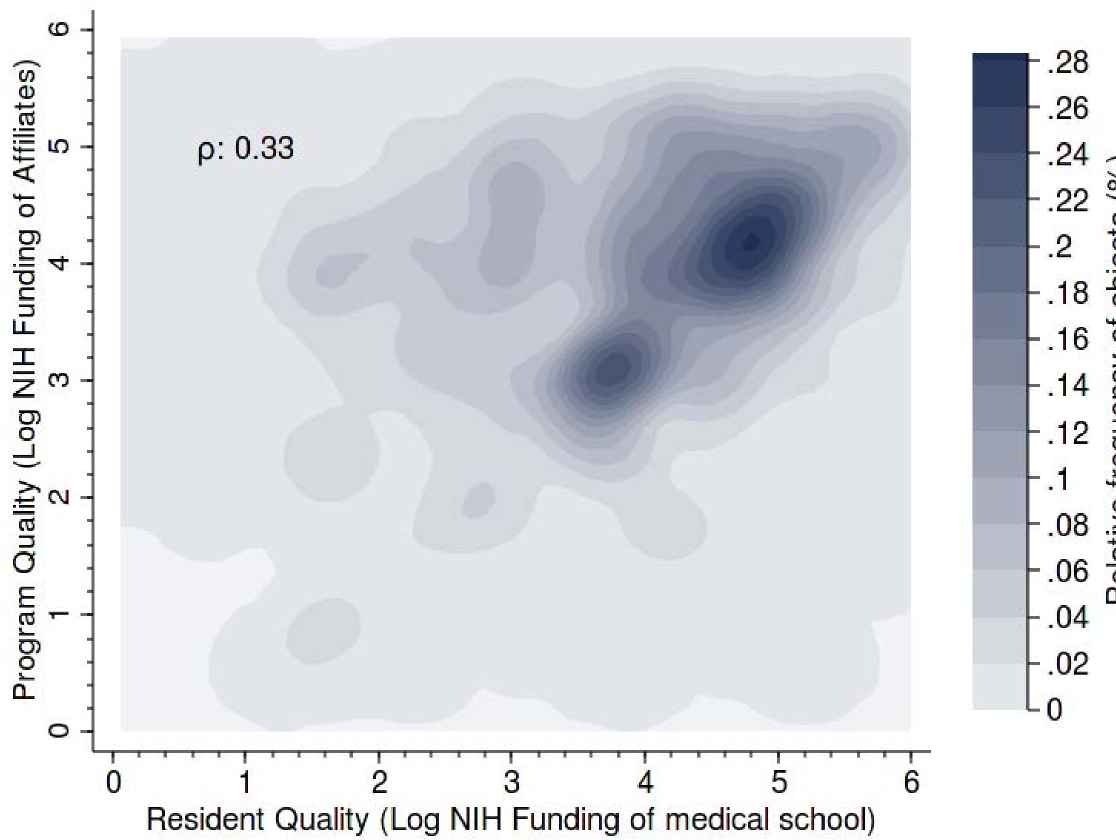
Notes: Linear Regressions. Dependent variable is the log of total implicit tuition at a residency program divided by the number of residents matched to the program. All regressions on generated implicit tuitions data using the 2010-2011 sample of residents and programs, and 100 simulation draws. All 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.

Table 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 (1)	\$10,000 (2)	\$20,000 (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

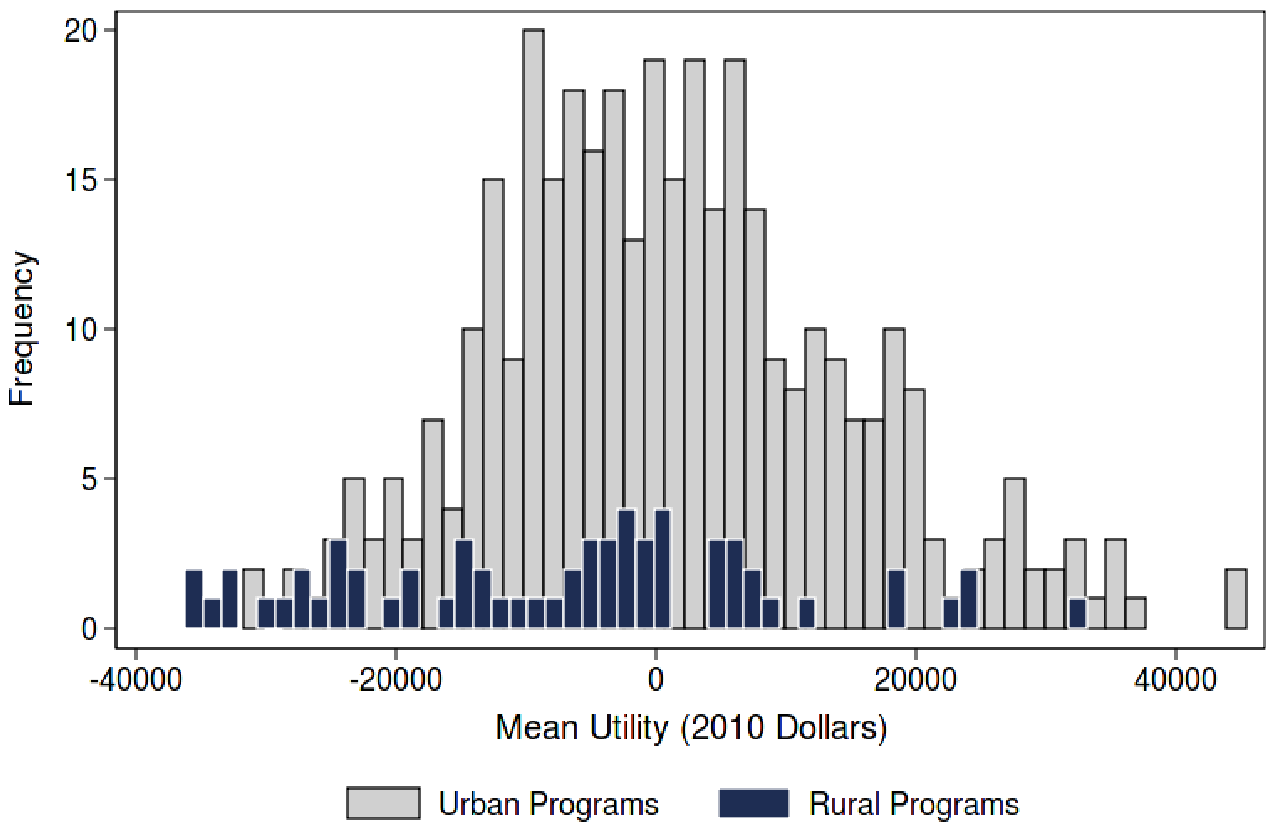
Notes: In Panel C, Column (i) decreases the urban positions in proportion to program size, subject to integer constraints. Positions at urban programs were reduced in proportion until further reductions would yield a greater number of residents than programs. In column (i), this yielded 32 more positions 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.

Figure 1: Assortative Matching between Programs and Residents



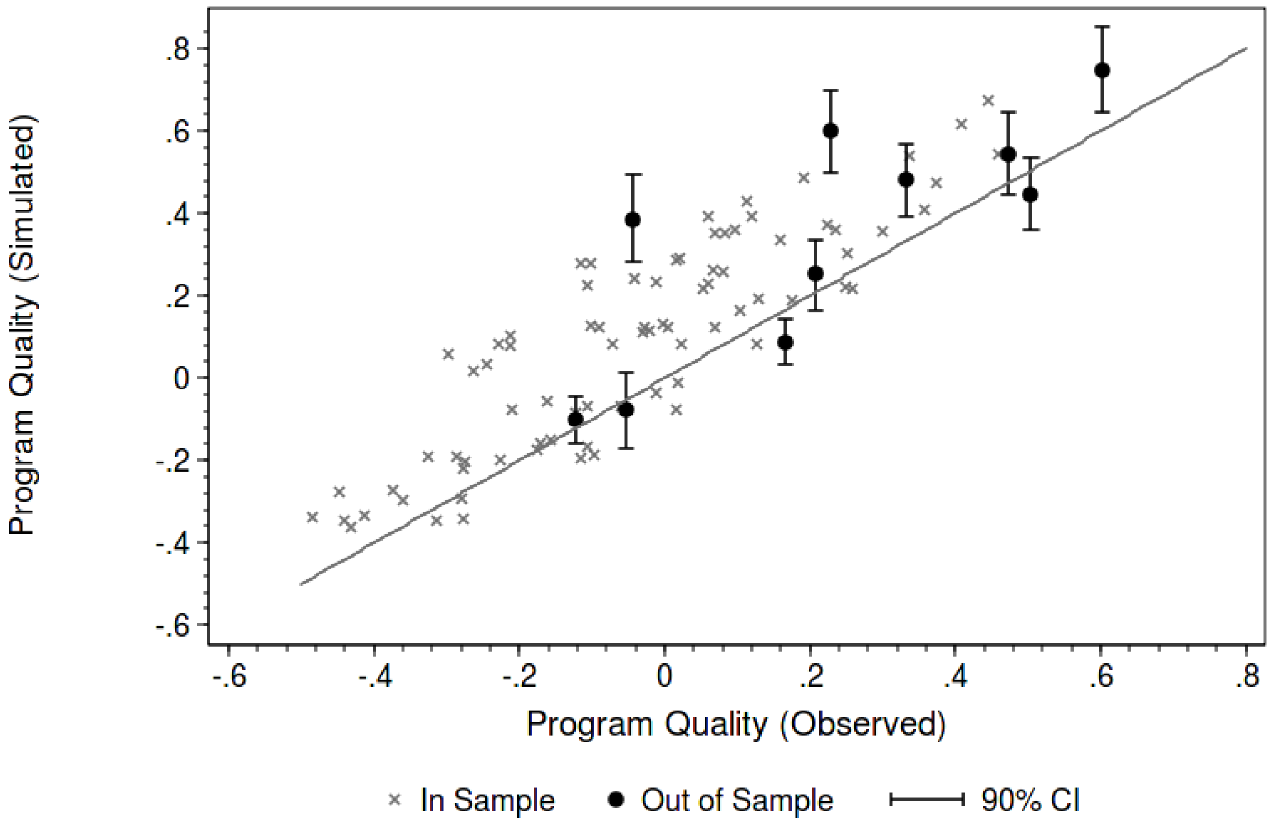
Notes: Darker regions depict higher density. Density calculated using two-dimensional bandwidths using a quartic kernel and a bandwidth of 0.6. Log NIH Fund of Affiliates is the log of the average of NIH funds at major and minor affiliates. Sample restricted academic year 2010-2011 and programs with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

Figure 2: Estimated Distribution of Program Utility



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



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.