Appendix for Online Publication
"An Empirical Model of the Medical Match"

Nikhil Agarwal*

December 27, 2014

The appendices follow the organization of the paper. Appendix A details the data sources and the procedures used to link and clean the data. Appendix B presents further details on the estimation procedure, Monte Carlo simulations to assess the sensitivity to mis-specification, and details on the optimization algorithms used in the paper. Appendix C presents details on the instrument. Appendix D discusses the parameter estimates under the alternative models. Appendix E presents proofs of the theoretical results on salary competition used in the paper.

A Data Construction

A.1 National GME Census

The American Medical Association (AMA) and the Association of American Medical Colleges (AAMC) jointly conduct an annual National Graduate Medical Education Census (GME Track) of all residency programs accredited by the Accreditation Council for Graduate Medical Education (ACGME). There are two main components of the census: the program survey and the sponsoring institution survey. The program survey, which is completed by the program directors, also gathers information about the residents training at the programs. Fields from the surveys are used to update FRIEDA Online, a publicly accessible database, and the AMA physician master-file. Since 2000, the GME Track has been pre-augmented with data from the Electronic Residency Application Service (ERAS) and the National Residency Matching Program (NRMP).[1] The AMA provided records from the National GME census on all family medicine residency training programs in the Unites States between 2003-2004 and 2010-2011. The 2011-2012 data was provided after the initial empirical analysis was completed.

The data files and identifiers are structured as follows:

1. Program file with program name, characteristics, a unique identifier for the program. This file also contains the identifier for the program's affiliated hospitals.

*MIT Department of Economics and NBER. Email: agarwaln@mit.edu.

[1]The details of the data collection procedure are outlined on http://www.ama-assn.org/ama/pub/education-careers/graduate-medical-education/freida-online/about-freida-online/national-gme-census.page.

2. Resident file with resident characteristics, program code, country code and medical school code. Two separate files identify the country and MD granting medical schools by name.

3. Institution file with the institution name, characteristics and a unique identifier.

4. Two bridge files. One delineating the relationships between programs and institutions (usually hospitals) as primary institution, sponsoring institution or clinical affiliate, and the other delineating the relationships between institutions and medical schools as major affiliate, graduate affiliate or limited affiliation.

A.1.1 Sample Construction

The baseline sample is constructed from the set of all family medicine residency programs accredited by the ACGME and first-year residents training at such programs. From this set, I exclude programs in Puerto Rico, military programs, and their first-year residents. Less than 20 programs and 123 residents are excluded due to these cuts. I also exclude programs that do not participate in the National Residency Matching Program and the residents matched to these program. These constitute less than 9 programs and 22 residents in each year. Finally, I also exclude the set of programs not offering any first-year positions, and programs that have no reported first-year matches during the entire sample period from the analysis. This final exclusion leads to 21 programs being dropped from the sample in 2003-2004, and less than 5 programs being dropped in the other years.

A detailed breakdown of the annual counts of the sample selection procedure is provided in Table A.1.

A.1.2 Merging GME Track Data

Programs to Clinical Site I wish to identify the primary hospital at which the clinical training of the residents in the program occurs. The AMA data identifies the relationship between programs and sponsoring institutions and hospitals in two ways. The program files records list each program's primary site. The program-institution bridge file records the sponsoring institution, (a second) primary clinical site and other affiliated institutions.

The program-institution bridge has the drawback that the clinical site of the program is not very well reported in the program-institution bridge with at most 94 observations (amongst all ACGME family medicine programs) in any given year whereas the sponsoring institutions are often medical schools or health systems. In order to avoid prioritizing sponsoring institutions or clinical sites from the bridge file, I pick the primary clinical site as reported in the program file as the starting point.

In a large number of cases, the institution type of the primary institution was a medical school or a health system, not a hospital. Consequently, the hospital institution data for these observations were not available. In the vast majority of these cases, the primary institution, at some point during the sample period was reported as a different site, one that was a hospital. I checked all cases in which the primary institution was not a hospital or clinic as identified by an institution type field in the institution file, or had a bed count of zero. When possible, I changed the primary hospital of a program from the listed program according to the following rules:

1. I first checked the program-institution bridge for a listed primary clinical site that was a hospital and changed the primary hospital to that primary clinical site.

2. I looked at the closest year in which the program listed a primary clinical site that is a hospital or clinic and changed it to that hospital or clinic only if the institution was listed as an affiliate or sponsor in that year as well.

The changes affected a total of 285 out of 3441 program-year primary clinical institution relationships in 109 out of 462 programs in the unrestricted sample of all family residency programs between 2003-2004 and 2010-2011. No more than 43 programs were affected in any given year.

Finally, 82 program-year observations did not have institution data from the primary sites based on the designation above. These programs were solely sponsored by health systems or medical schools, and not primarily associated with a hospital. I imputed the hospital characteristics by taking the mean characteristics of all hospital affiliates for these programs. This imputation populated records in 11 programs in 2003-2004 and 2004-2005 and 10 records in the other years.

Programs with Medical Schools The link between medical schools and programs is provided by the AMA through the program-institution bridge followed by the institution-medical school bridge. The program-institution relationships are categorized into primary clinical sites, sponsors and affiliates. The institution-medical school relationships are categorized as limited, graduate and major.

I use these relationships to define two types of affiliations for programs to medical schools, major and minor. A program has a major affiliation to a medical school if the primary or sponsoring institution has a major affiliation with a medical school. All other relationships are regarded as minor relationships. The relationships between programs and medical schools are imputed for all years between the first and last year of a major (likewise minor) relationship. I used all relationships since 1996 for this imputation. For the unselected sample of family medicine programs between 2002-2003 and 2010-2011, I imputed relationships for 144 out of 2797 major affiliations and 702 out of 3337 minor affiliations. The mean NIH funding across all major and minor affiliations are used as the variables for this merge.

A.2 Medical School Characteristics

The National GME Census does not provide data on medical school characteristics. Each medical school is identified by a number, and only the medical school names for MD granting medical schools are identified. According to the AAMC, there are 134 accredited MD-granting medical schools in the United States. In the dataset, I found 135 medical school identifiers for MD granting institutions. Texas Tech University Health Sciences Center School of Medicine appeared with two different ids. I duplicate the fields throughout for that medical school. I next describe the sources of the data on medical schools and the process used to merge and construct the fields.

A.2.1 NIH Funding Data

The National Institutes of Health organizes the data on its expenditures and makes it available through RePORT. The records of each project funded by the NIH is available for download through http://projectreporter.nih.gov/reporter.cfm. The records identify the projects by an application id and fields include the institution type, total cost and project categories. I included funding for projects designated to Schools of Medicine, Schools of Medicine and Dentistry, and Overall Medical as these categories were the major categories at which the recipient was affiliated with an MD medical school. I include funding only for extramural and cooperative research activities, and training and fellowship programs funded by the NIH in a medical school. So, I dropped activity codes beginning with G, C, H as these were designated for construction, resource development and community service. Further, I dropped activity codes beginning with N and Z since those data are available only after 2007.

I used the records from all project costs incurred in the financial years 2000 to 2010 that satisfy the criteria above and aggregated the project costs to the organization name. I constructed the average annual NIH research costs incurred at these medical schools during this period. I infer that a school was operating during a given year if it secured some NIH funding. All but thirteen schools secured NIH funding during each of the eleven years in the sample. Six schools did not receive any NIH funds during this period even though they were operating (as indicated by online sources) and their eleven year annual average NIH costs were set to zero. For the remaining seven medical schools, I established the number of years the school was operating by searching for the history of the school from the history of the medical college published on their websites.

These data were merged with the data from the National GME Census using the medical school names. Of the 135 MD medical schools in the GME Census, 129 medical schools were matched successfully to a counterpart in the NIH funding data. I verified that the remaining six schools did not have any records in project RePORT in the categories considered.

A.2.2 Medical School Admission Requirements (MSAR)

I used the records from the 2010-2011 MSAR publication of the AAMC to augment the medical school characteristics with the state and the median MCAT score of the admits into a medical school. The merge was done using the medical school name and MCAT score data was found for all but seven of the 135 MD granting medical schools. Data on the state the medical school is located in was found for all MD medical schools.

A.3 Medicare Data

Here, I describe the merge and construction of the Case Mix Index and Wage Index variables. The instrument, based on Medicare reimbursement rates is described in Section C.

I used the records from the Medicare provider files to construct the variables primary care reimbursement rates, the Medicare wage index and the case mix index. The institution ids for all affiliates were merged with Medicare provider identifiers by the name of the provider by using the 1997 PPS files, and then using the 2010 Impact Files. A second check was conducted for primary institutions of the programs, and for affiliates when primary institutions were not matched to Medicare data. In a small number of instances, there are multiple matched CMS identifiers for a single institution. Medicare variables were averaged across these multiple matches.

A.3.1 Medicare Wage Index and Case Mix Index

The Center for Medicare Services calculates a Wage Index and Case Mix Index for each provider.[2] I merged the CMS data with primary institution. In a small number of instances, the primary institution did not have a match with Medicare data. In these cases, I calculated the average of the variable for all affiliates with Medicare data. In 63 out of 3441 cases, the case mix index was not available even for affiliates. Here, in the structural estimates, I used an imputed value from a linear regression on all other characteristics included in the demand system. Finally, missing values of the wage index were imputed using the geographic definitions Medicare uses to calculate the wage index.

A.4 Identifying Rural Programs

I use two sources of data to identify the set of rural family medicine program.

1. The American Academy of Family Physicians has a program directory of all family medicine programs in the United States. The program directory lists the community setting of the program as one or more of Urban, Suburban, Rural, Inner-city. Programs for which only rural was listed as the community setting are considered rural programs by this definition. The records from this directory were scraped on 01/05/2012. I manually merged the set of rural programs to AMA data using the name of the program, the hospital and the street address. In the years 2003-2004 to 2010-2011, this procedure identified 438 program-year observations as rural programs.

2. The program names in the AMA data often directly indicate whether a program is a rural program or not. For instance, the University of Wisconsin sponsors several programs in family medicine, one of which is named "University of Wisconsin (Madison) Program" and

[2]The files and the description of the calculation for the wage index is given on http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/wageindex.html and the Case Mix Index is described on http://en.wikipedia.org/wiki/Case_mix_index

the other named “University of Wisconsin (Baraboo) Rural Program.” I consider all programs with “Rural” in the name during the same period of the program as a rural program. This procedure identified a total of 159 program-year observations as rural programs in the years 2003-2004 to 2010-2011, of which a total of 115 program-year observations overlapped with program-year observations identified as a rural program using the previous procedure.

Using the 2010-2011 data, I checked for contradictions where a program with rural in the program name listed a community setting other than rural in the AAFP directory. There were a total of 5 programs that were classified as rural according to rule 2 but not rule 1. Of these, in four cases, the program directory did not have any information other than the name and address of the program. The community setting for the remaining program was listed as suburban as well as rural.

A.5 Resident Birth Location

The birth location of the resident is recorded as city, state and country code. The following steps were carried out to improve the quality of the data and then to identify whether a resident was born in a rural location in the United States:

1. I convert the AMA country identifiers, which are not unique across years, to the corresponding ISO 3166-1 alpha-3 identifier using the country name provided by the AMA. Except for some former soviet nations and territories of the UK, US and Netherlands, a unique match was available.

2. The state and country for observations with only the city name were imputed using the state and country for an identically spelled city if that state-country combination constituted more than 50% of the observations for that city. This imputation was carried out using the GME Census data from 1996-1997 to 2010-2011 in five specialties: internal medicine, pediatrics, OB/GYN, pathology and family medicine.

3. For US born residents, city-state combinations were geocoded. The observations for which the geocoder indicated a match with unexpected accuracy (more than, or less than city level accuracy) were checked by hand and minor spelling errors were corrected. The corrections were put through the geocoder for a second time. Ambiguous entries were coded as missing data.

4. The county of birth for US born residents was extracted matched with a list of counties that belong to a Metropolitan Statistical Area in order to construct the rural birth indicator.

A.6 Other Data

A.6.1 CPI-U

I downloaded the records of the monthly Consumer Price Index for All Urban Consumers from the Federal Reserve Economic Data (FRED) website. I use the December observation for the CPI-U for a year.

A.6.2 Rent

Data from the 2000 US Census was downloaded from nhgis.org. I used county level aggregates from sample file 1 for population, age and race variables, and from sample file 2 for income and rent variables. The median gross rent is used as the measure of rent as it adjusts for the utility payments.

The 2010 US Census did not use the long form on which data on the rent paid is collected. Consequently, data on the county level median gross rent was downloaded from the 2006-2010 American Community Survey using Social Explorer. These rent numbers are adjusted to 2010 dollars by Social Explorer. The five-year aggregate was preferred to the annual or three-year aggregates since the latter did not cover all counties in the US.

To construct the median gross rent variable, I convert the median rent data from the 2000 US Census into 2010 dollars by using CPI-U. A linear interpolation between the 2000 and 2010 rent data for the interim years.

Merging The city, state and zip code of the program and institutions were used to geocode the latitude and longitude of the zip code's centroid. These latitudes and longitudes were then used to determine the county in which the program or institution is located using county shape files provided by NHGIS. The geographic ids from this process were used to merge these with the data files. Every program in the sample was successfully matched in this process.

A.7 Miscellaneous Issues

1. For the preference estimates, imputation of salaries for missing data was done for 23 observations out of 3,441 using a linear regression on the other characteristics included in the model.

2. The program survey asks for the number of first year positions offered in the next academic year. I use this as the preferred measure of the program's capacity when available. In ten instances, this field was not available and for nine of these instances, it was imputed from the value of the field from the previous year. In the remaining instances, the number of first year residents in the program was taken to be the number of positions offered. I checked that the reported number positions offered next year is typically equal to value of the field from the previous year.

I found some instances when the number of residents in first year positions exceeds this capacity measure. In these cases, I take the maximum of the number matched to the program and the lagged response to the first year enrollment as the program's capacity. In more than 75% of the cases, the number matched did not exceed the reported number of positions by more than one. Table A.2 summarizes the number of observations affected by this change and the mean size of the change. One reason for the discrepancy may be residents that repeated their first year training or deferred enrollment.

B Estimation Appendix

B.1 Sensitivity of Estimates to Mis-specification

This section presents Monte Carlo results aimed at assessing the sensitivity of the estimator used in the paper to model mis-specification. Perhaps the two strongest assumptions in the paper are that (1) preferences on one side of the market is homogeneous and (2) observed matches are pairwise stable. I assess the sensitivity of the estimation procedure by simulating data-sets that violate these assumptions and estimate a model that specifies preferences for residents and programs as in Section III.A:

$$\begin{aligned} h_i &= x_i\alpha + \varepsilon_i \\ u_j &= z_j\beta + \xi_j \end{aligned} \tag{B.1}$$

where x_i and z_j are observables, and ξ_j and ε_i are unobserved. All characteristics are drawn standard normal random variables, distributed independently of the other variables.

I focus on simulations with only one characteristic for computational simplicity given the large number of optimizations involved in this exercise and the primary concern that mis-specification may reflect in the unobservable characteristics. As in other empirical exercises, deviations from equilibrium assumptions that are systematically related to observable agent characteristics, or correlations of unobservables with observables are likely to yield biased estimates.

Ideally, one would use the observed characteristics in the simulations so that the properties of the estimator for this empirical exercise could be better evaluated. Doing this would require repeatedly optimizing over several parameters, increasing the computational costs of considering many variations of the baseline model. Instead, my approach is to simplify the computation in order to consider different forms of mis-specification and to evaluate sensitivity of estimates to the degree of mis-specification. The results should be interpreted with some caution given these differences.

To mimic the size of the data-set used in this paper, I simulated data from markets in which there are 400 programs and 2,800 residents. I endow each program with 7 positions. The average year in the data-set for the study has 428 programs and 3,148 residents respectively.

For the Monte Carlo simulation exercise, I use the overall sorting moments and the within-program variance as described in equations (11) and (12) of the main text. The moment in equation (13) of the main text is not relevant for the exercise since each resident has only one characteristic.

The Monte Carlo exercises use $B = 1000$ simulated data-sets under varying types of mis-specification. For each such data-set, indexed by b, I estimate $\hat{\alpha}^b$ and $\hat{\beta}^b$ by finding the parameter that minimizes the Euclidean distance between the moments in the generated data-sets and the model's predictions. I use a grid-search with α and β on $\{0, 0.01, \ldots, 2\}$ and a baseline model with true parameter value of $\alpha = \beta = 1$.

In addition to standard Bias and RSME of the parameters, I examine at the Pseudo-R^2 statistic for estimated α and β:

$$\begin{aligned}\text{Pseudo } R^2_\alpha &= \frac{\hat{\alpha}^2 Var(X)}{\hat{\alpha}^2 Var(X) + Var(\varepsilon)} \\ \text{Pseudo } R^2_\beta &= \frac{\hat{\beta}^2 Var(Z)}{\hat{\beta}^2 Var(Z) + Var(\varepsilon)}.\end{aligned} \tag{B.2}$$

These quantities measure the fraction of the variance in the latent utility that is explained by the observables and are an important statistic of the preference distribution. For example, if z is the salary of the program, the Pseudo R^2 measures how much of the differences across programs is captured by the salary, and is therefore important for counterfactuals that depend on the importance of salaries as a determinant of preferences.

The remaining subsections discuss Monte Carlo results from the types of mis-specification described earlier. Section B.1.1 assesses biases from mis-specified unobserved heterogeneity, Section B.1.2 considers an environment with interview frictions and Section B.1.3 considers robustness to random deviations from pairwise stability.

B.1.1 Preference Heterogeneity

One significant concern is that the model does not allow for match-specific idiosyncratic preferences, which may result in significantly biased estimates of α and β. To assess the sensitivity of estimates, I simulated the resident optimal pairwise stable match for economies in which preferences are specified as

$$\begin{aligned}v_{ji} &= x_i\alpha + \varepsilon_i + \delta_\alpha \eta_{ji} \\ u_{ij} &= z_j\beta + \xi_j + \delta_\beta \nu_{ij}\end{aligned}$$

where $x_i, z_j, \varepsilon_i, \xi_j$ are as in equation (B.1) above, η_{ji} and ν_{ij} are standard normal random variables, and α and β are set to 1. Here, v_{ji} represents the latent value of program j for resident i and u_{ij} represents the latent utility of resident i for program j. I can control the importance of unobserved idiosyncratic preferences by varying the values of δ_α and δ_β. I then estimated a (mis-specified) model that assumes that $\delta_\beta = \delta_\alpha = 0$. Hence, the values of δ_β and δ_α control the degree of mis-specification.

Table B.1 presents the Monte Carlo results from 1,000 simulation exercises. The first row, with $\delta_\alpha = \delta_\beta = 0$ provides estimates from the correctly specified model. The Bias and RMSE are orders of magnitude lower than the true values of α and β. The remaining rows increase the degree of mis-specification using the parameters δ_α and δ_β. The results from the top panel indicates that as δ_α is increased, the magnitude of the bias in α increases, but there is little effect of this form of mis-specification on the estimates of β. Remarkably, the robustness of estimates of β appear to hold even if δ_β is increased (although the root-mean-squared error for β increases slightly). For large values of δ_α and δ_β, we do see significant bias in the estimates of α. Hence, it appears that the estimated average utility of residents is robust even to a large degree of mis-specification of this type. On the other hand, the programs' preferences for residents are more sensitive. Nonetheless, the bias is less than 10% for values of δ_β and δ_α of up to 0.5 (up to 50% of the vertical unobservables).

From the perspective of estimating the relative importance of observables, however, the results indicate that large values of δ_α and δ_β typically result in an over-estimated importance of observables for preferences on both sides of the market.[3] This is because the estimated model does not correctly account for all sources of residual variance. Nonetheless, for moderate values of δ_α and δ_β (up to 50% of the vertical unobservables), the bias in the Pseudo-R^2 measures is less than 6 percentage points. This observation is comforting for the robustness of the method to small degrees of unobserved idiosyncratic preferences.

B.1.2 Frictions from not Interviewing all Applications

Instead of incorrectly specified preferences, one may be concerned that the observed matches are not exactly pairwise stable. Particularly, the medical residency market and many other matching markets may have an interview stage in which it is possible that some agents do not interview with their stable match partners. These interviewing frictions could result in biased preference estimates even if the underlying preferences are correctly specified.

I use a stylized model of interviewing frictions to assess the robustness of the estimation method. The model for interviewing frictions assumes that a fraction ρ of randomly chosen pairs of agents do not meet with each other and therefore cannot match. The preferences over a partner that an agent has met with are as in equation (B.1) above, and the (unique) match with no blocking pairs is observed. Larger values of ρ indicate larger interviewing frictions. For large values of ρ, we expect some agents to not be matched, and I assume that such agents are missing from the data-set. I then simulate data-sets generated by this process and estimate a model that does not account for any interviewing friction.

[3] For this model, the true Pseudo R^2 is given by

$$\begin{aligned}
\text{Pseudo } R^2_\alpha &= \frac{\hat{\alpha}^2 Var(X)}{\hat{\alpha}^2 Var(X) + Var(\varepsilon) + \delta_\alpha^2 Var(\eta)} \\
\text{Pseudo } R^2_\beta &= \frac{\hat{\beta}^2 Var(Z)}{\hat{\beta}^2 Var(Z) + Var(\varepsilon) + \delta_\beta^2 Var(\nu)}
\end{aligned}$$

while the model estimated values are given by equation (B.2) above.

Table B.2 presents the Monte Carlo results from 1,000 simulation exercises. The estimates are remarkably robust to these frictions with a small bias and root mean squared error for both parameters as well as the related Pseudo R^2 statistics. Further, when $\rho = 0.4$, on average, only 2.3 out of 2,800 residents are not matched. These observations indicate that this particular nature of interviewing frictions results in agents matching with observably very similar programs for the chosen parameters. The results suggest that the methods may not be significantly biased due to this form of mis-specification. It is important to note, however, that the nature and degree of interviewing frictions in the real world may differ from the particular form and values of ρ chosen in the simulations.

B.1.3 Random Permutation of a Fraction of Matches

The simulations of interviewing frictions picks a particular form of mis-specification from pairwise stability that may be different from the predominant deviations in any given setting. As it is infeasible to simulate outcomes from every possible model, I consider a final stylized form of mis-specification in which a fraction κ of residents are randomly permuted after a pairwise stable match is computed from the model as described in equation (B.1) above.

Table B.3 presents the Monte Carlo results from 1,000 simulation exercises. The main effect of this type of mis-specification is a negative bias in the estimate of α. This is not very surprising since shuffling the residents without respect to preferences of the programs will result in a larger within-program variance in resident characteristic. Interestingly, the effect on the estimate of β is not as pronounced. The bias and the Pseudo-R^2 statistic for the resident preferences are within 1.2% even for $\kappa = 0.4$. This is comforting for robustness of counterfactuals that are based on estimates of residents' preferences alone.

B.2 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, modeling the data generating process as independently sampled matches is unappealing as well.

Agarwal and Diamond (2014) consider a data generating process in which the number of programs and residents increases and each program has two positions. The observed data is a pairwise stable match for N residents and J programs with characteristics (x_i, ε_i) and (z_{jt}, ξ_{jt}) drawn from their respective population distributions. These large market asymptotics are appealing in this setting since the family medicine residency market has about 430 programs and 3,000 residents participating each year. The challenge in obtaining asymptotic theory arises precisely from the dependence of matches on the entire sample of observed characteristics. They prove that an estimator that minimizes the distance between sample moments and those predicted as a function of θ is consistent for the double-vertical model in a single market. They also present Monte Carlo evidence on a simulation based estimator for a more general model like the one estimated in this. Simulations suggest that the root mean square error in parameter estimates decreases with the sample size.

Motivated by Agarwal and Diamond (2014), I compute the covariance of the moments is estimated using a parametric bootstrap to account for the dependence of matches across residents and approximate the error in the estimated parameter using a delta method that is commonly used in simulated estimators (Gourieroux and Monfort, 1997):

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

where $\hat{\Gamma}$ is an estimate of gradient of the moments with respect to θ evaluated at $\hat{\theta}_{SMD}$; W is the weight matrix used in estimation; $\hat{V}$ is a bootstrap estimate of the covariance of the moments at $\hat{\theta}_{SMD}$; S is the number of simulations and $\hat{V}^S$ is an estimate of the simulation error in the moments at $\hat{\theta}_{SMD}$.

To estimate the derivatives, I construct two-sided numerical derivatives of the simulated moment function $\hat{m}(\theta)$ using the observed population of residents and programs. Since $\hat{m}^S(\theta)$ is not smooth due to simulation errors, I use 10,000 simulation draws and a step size of 10^{-3}. The simulation variance is estimated by calculating the variance in 10,000 evaluations of $\hat{m}^S\left(\hat{\theta}_{SMD}\right)$, each with a single simulation draw and using the observed sample of resident and program characteristics. These two quantities can be calculated independently in each of the markets.

B.2.1 A Bootstrap

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

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

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

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

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

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

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

The bootstrap replaces the population distribution of observed characteristics of the residents and programs with the empirical distribution observed in the data. Given a sampled economy, it computes $\hat{\nu}_{jt}$ and the moments at a pairwise stable match at $\hat{\theta}$. The covariance of the moments across bootstrap iterations is the estimate of $\hat{V}$. The uncertainty due to simulation error $\hat{V}^S$ is approximated by drawing just the unobserved characteristics from the assumed parametric distribution.

The method yields consistent estimates for standard errors if the equilibrium map from θ and the distribution market participants to the data is smooth. Standard Donsker theorems apply for the sampling process for market participants. The inference method above should then be consistent if a functional delta method applies to this map i.e. the distribution of the moments is (Hadamard) differentiable jointly in the parameter θ and the distribution of observed characteristics of market participants.

B.2.2 Weight Matrix

In Monte Carlo simulations with this dataset and I found that using the identity matrix was often inaccurate and left us with a poor estimate of θ_0. Intuitively, the identity matrix fails to account for the co-variance in the various program and resident characteristics as well as the covariance with the within-program moments. To appropriately weight some of these aspects, I use a weight matrix $\tilde{W}$ that is calculated using the following bootstrap procedure. For each market t, with replacement, randomly sample J_t programs and the residents matched with them. Treat the observed matches as the matches in the bootstrap sample.[4] Compute moments $\left\{\tilde{m}^b\right\}_{b=1}^{B}$ from the sample and compute the variance $\tilde{V}$ and set $\tilde{W} = \tilde{V}^{-1}$. While this weight matrix need not converge to the optimal weight matrix, the only theoretical loss is in the efficiency of the estimator. This weight matrix also turns out to be close to one that would be calculated as $\hat{W} = \left(\tilde{V}\left(\hat{\theta}_{SMD}\right) + \frac{1}{S}\hat{V}^S\left(\hat{\theta}_{SMD}\right)\right)^{-1}$ where $\hat{\theta}_{SMD}$ is the estimate of θ_0 using $\tilde{W}$ as the weight matrix.

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

B.3 Optimization Algorithm

The function defined in equation (10) may be non-convex and may have local minima. Further, since $\hat{m}^S(\theta)$ is not smooth as it is simulated. Gradient based global search methods can perform very poorly in such settings. I use an extensive derivative free global search followed by a refinement step that uses a derivative free local search to compute the estimate $\hat{\theta}_{SMD}$.

The global search is implemented using MATLAB's genetic algorithm and a bounded parameter space based on initial runs (Goldberg, 1989). As with all algorithms on non-convex problems, there is no guarantee that the genetic algorithm finds the global optimum. I conducted three initial genetic algorithm runs to with separately seeded populations of size 40, cross-over fraction of 0.75, one elite child, an adaptive mutation scale of 4 and shrinkage of 0.25. These extensive runs were used to generate starting values for the local searches.

Local searches using these starting values and from minimizers of similar models were implemented. The step is conducted to refine the estimate $\hat{\theta}_{SMD}$ and to be thorough in the search for the global minimum. I used the NLOpt (Johnson, 2011) implementation of the subplex algorithm (Rowan, 1990), a variant of the Nelder-Meade algorithm, also a derivative free optimization routine. Since there is a possibility that the algorithm gets stuck in local minima, I use up to three successive runs of the subplex algorithm. Each run restarts the algorithm using the optimum found in the previous run. I do not repeat the local search if the change the point estimate between the starting value and the optimum is less than 10^{-6} in Euclidean norm. Two iterations were always sufficient and the parameter value in the local run was typically close to the output from the genetic algorithm. I also verified that the reported point is a local minimum on one dimensional slices of the parameter space by profiling the objective function.

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

C Medicare Reimbursement Rates and Instrument Details

C.1 Description of Medicare Reimbursement Regulations

Medicare Direct Graduate Medical Expenditure (DGME) payments are designed to compensate teaching hospitals for expenses directly incurred due to the training of residents. The methodology used to determine these payments was established in the Consolidated Omnibus Budget Reconciliation Act (COBRA) of 1985, and are implemented as per 42 CFR §§ 413.75 to 413.83. Here, I provide a broad outline of the method used to determine Medicare DGME payments and the PCPRA variable used in the analysis.

Roughly, the total DGME reimbursements to a hospital is the product of the hospital specific per resident amount (PRA), the weighted number of full-time equivalent residents (FTE) and Medicare's share of total inpatient days. The PRA is determined using the total costs of salaries and fringe benefits of residents, faculty and administrative staff of the residency program and allocated institutional overhead costs divided by the total number of full time equivalent residents in a base year, usually 1984 or 1985. Hospitals that began sponsoring residency training after 1985 were grandfathered into the program using their first year of reported costs as the base year. After 1997, a new hospital's per resident amount was based on the reported costs of other programs in the geographic area, which is an MSA/NECMA, rest of state or a census division depending on the number of other providers sponsoring GME. The Balanced Budget Act of 1997 also introduced certain ceilings and floors on the per resident amount. See Gentile Jr. and Buckley (2009) for a more comprehensive legislative history of Medicare reimbursement of Graduate Medical Education.

Between 1985 and 2000, the PRA for a hospital was revised by adjusting for the 12 month change in CPI-U, and minor changes on previously misallocated costs. An exception was made in 1993 and 1994 when two separate PRAs were effectively created, one for primary care and obstetrics and gynecology residents and the other for all other residents. In these two years, the non-primary care PRA was not adjusted for inflation.

Subsequent to 2000, the per resident amounts were also adjusted using the change in CPI-U but were subject to a floor and ceiling put in place by the The Balanced Budget Act of 1997. The floor increased the PRAs of hospitals that were below 70% of the (locally-adjusted) national average per-resident amount to 70% of the total and later to 85%. The ceiling gradually decreased the PRAs of hospitals that were above 140% of the (locally-adjusted) national average per-resident amount until the PRA of a hospital fell below the ceiling. The exact procedure used to make these adjustments is detailed in 42 CFR § 413.77. The Balanced Budget Act of 1997 also created new regulations on the manner in which the number of full-time equivalent residents was determined. These regulations are detailed in 42 CFR § 413.86.

C.2 The Instrument: Competitor Reimbursement Rates

To construct competitor reimbursements, I first extract the records from the fields "Updated per resident amount for OB/GYN and primary care" and "Number of FTE residents for OB/GYN and primary care" on lines 2 and 1 respectively in form CMS-2552-96, Worksheet E-3, Part-IV for the cost reporting period beginning October 1, 1996 and before September 30,1997. As per the instructions for this form (3633.4), this is the last period for which the response to the field was required by the hospitals. Indeed, I found only five observations for this field in the cost reporting period ending October 1, 1998 and no observations in the next period. The per resident amount variable is recorded in cents, and so is first converted into dollars. Both fields were winsorized at the bottom at top 1 percent since the range of values were extreme. Barring the effects of winsorizing the data, the distribution of the per resident amount variable is similar to Exhibit 2 in Newhouse and Wilensky (2001). While some institutions have per resident amounts less than $40,000, others are reimbursed at rates higher than $200,000.

The Competitor Reimbursement variable for an institution is constructed in order to mimic the per resident amount calculation done by Medicare for new sponsors. As given in equation (7), the (weighted) Competitor Reimbursement variable for a program is the average (weighted by FTE) of all primary care per resident amounts in the primary institution's geographic area (MSA/NECMA or the rest of the state) other than that of the primary institution. When this average is constructed from less than three observations, the census division is used. This variable is then merged to the primary institution of a program as defined earlier.

Figure C.3 depicts the state-averaged variation in the instrument that is not explained by the controls included in the preference estimates and a program's own reimbursement rate. A degree of spatial correlation within a census division is noticeable due to the definition of the geographical units used. To assess whether the cross-sectional nature of the instrument is likely to result in biased estimates, Table C.1 presents regressions of the instrument on characteristics included in the preference estimation, as well as excluded location characteristics such as median age, median household income, crime rates, total population and college share. The instrument yields biased results if the excluded location characteristics are strongly correlated with the instrument and affect preferences. Several pieces of evidence suggests that any bias is likely to be small. First, estimates in Columns (2) and (4) indicate that the coefficients on most of the excluded characteristics are not statistically significant predictors of reimbursement rates or the instrument. Median age in the county and property crime rates are exceptions, but the economic magnitudes of the coefficients on these two variables are small. Estimates in Column (2) imply that a one standard deviation change in the median age or property crime rates results in less that 4% change (0.2 standard deviations) in competitor reimbursement rates. Second, the excluded characteristics together explain less than 6% of the variation not explained by the other variables that are included in the preference model. Columns (3-4) show that characteristics of the program itself explain about 35% of the variation in its reimbursement rates and the addition of location characteristics is not important. Finally, note that Column (6) in Table 6 indicates that the relationship between the instrument and the endogenous variable is not driven by these characteristics. Together, these observations suggests that any bias due to the instrument is likely to be small as the included controls are sufficient. These findings are consistent with Anderson (1996), which argues against this reimbursement schemes on the basis that other cost predictors do not correlate very strongly with per resident amounts.

D Parameter Estimates

Table D.1 presents point estimates of the models discussed in Section VI. Panel A presents parameter estimates for the distribution of residents' preferences and Panel B presents estimates for the human capital index. As mentioned in the text, these point estimates are not directly interpretable in economically meaningful terms. Table 7 translates a subset of coefficients from Panel A into monetized values by dividing a given coefficient by the coefficient on salaries, and scaling them into dollar equivalents for a one standard deviation change.

First, comparing coefficients on salaries from Specifications (1) and (2) to the corresponding Specifications (3) and (4), we see that accounting for endogeneity in salaries reduces the point estimate on the salary coefficient. Many of the other coefficients are not substantially altered by the inclusion of the control variable and the program's own reimbursement rates. The annual rent and NIH funding of major affiliates are two exceptions. This may be a consequence of correlation between reimbursement rates and these covariates. The primary economic implication of the drop in coefficient in salaries on including the instrument is that the willingness to pay for programs increases substantially.

Comparing estimates from various specifications, we also see changes in the estimated coefficient on the medicare wage index and rent. A reason for the change in the wage index and rent is that these variables are highly correlated with each other and with reimbursement rates. Across specifications, the relative magnitude on coefficients on rural birth interacted with rural program, program location in birth state and program location in medical school state have similar relative magnitudes although large in overall magnitude in Specification (2). I attribute this difference to additional unobserved heterogeneity in Specification (2), due to which similar geographic sorting needs to be explained with higher preference for these characteristics.

E Salary Competition

E.1 Expressions for Competitive Outcomes

I first characterize the competitive equilibria of the model. The expression in equation (14) follows as a corollary. For clarity, I refer to the quality of program 1 as q_1 although I normalize it to 0 in the model presented in the text.

Proposition E.1 *The salary w_k paid to resident k by program k in a competitive equilibrium is characterized by*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Corollary E.2 *The worker optimal competitive equilibrium salaries are given by*

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

and the firm optimal competitive equilibrium salaries are given by

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

E.2 Proof of Proposition 2

Consider an N-vector of outputs $y = (y_1, \ldots, y_k)$ and define a family of production functions $\mathcal{F}(y) = \{f : f(h_k, q_k) = y_k\}$ where y_k denotes the output produced by the pair $(h_k,\ q_k)$. The two extremal technologies in this family are given by $\bar{f}_y(h_k, q_l) = y_k$ and $\underline{f_y}(h_l, q_k) = y_k$ for all $l \in \{1, \ldots, N\}$. Let $w_k^{fo}(f)$ (likewise $w_k^{wo}(f)$) denote the firm-optimal (worker-optimal) competitive salary under technology f.

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

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

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

Proof. I only derive the bounds for the worker optimal equilibrium since the calculation for the firm optimal equilibrium is analogous. From the expressions in Corollary E.2,

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

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

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

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

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

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

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

■

Proposition 2 follows as a corollary:

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

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

■

E.3 Implicit Tuition

I prove a more general result for many-to-one assignment games that subsumes Proposition 3. A many to one assignment game between workers $i \in \{1, \ldots, N\}$ and firms $j \in \{1, \ldots, J\}$ is defined by the capacity of firms firms c_j and the surplus a_{ij} produced by the worker-firm pair (i, j). The surplus from multiple workers is additively separable and an empty position produces 0. I focus on the case when $\sum_j c_j \geq N$. I micro-found the surplus as the sum, $a_{ij}^f = u_{ij} + f(h_i)$, of the production $f(h_i) \geq 0$ produced by a worker with human capital h_i and the utility worker i receives from working at firm j at a salary of w, given $u_{ij} + w$. I assume that each $u_{ij} \geq 0$. For completeness, I define a few concepts below. Rigorous treatments of these concepts for the one-to-one case are given in Roth and Sotomayor (1992), in Camina (2006) and Sotomayor (1999) for the many-to-one case.

An **assignment** is a vector $x = \{x_{ij}\}_{i,j}$ where $x_{ij} = \{0, 1\}$ and $x_{ij} = 1$ denotes that i is assigned to j. The assignment x is **feasible** if $\sum_i x_{ij} \leq 1$ and $\sum_j x_{ij} \leq c_j$. In the many-to-one case, we refer to an **assignment of positions** $\{y_{i,p}\}_{i,p}$ where $p \in \left\{1, \ldots, \sum_j c_j\right\}$ denotes a position. Let j_p denote the firm offering position p. Each assignment x induces a unique canonical assignment of positions y where the positions in the firm are filled by residents in order of their index i.

An **allocation** is the pair (y, w) of an assignment of positions y and salaries $w = \{w_{ij}\}_{ij}$ with $w_{ij} \in \mathbb{R}$. The surplus of position p is defined as $v_p^f = \sum y_{ip}(f(h_i) - w_{ip})$ and of worker i by $u_i^f = \sum y_{ip}\left(u_{ij_p} + w_{ip}\right)$. An **outcome** is a pair $((u, v); y)$ of payoffs $u = \{u_i\}_i$ and $v = \{v_p\}_p$ and an assignment of positions y.

A feasible outcome $((u, v); y)$ is **stable** if for all i, p, we have $u_i \geq 0$, $v_p \geq 0$, $u_i + v_p \geq a_{ij_p}$ if $y_{ip} = 1$ or $x_{ij_p} = 0$, where x is the assignment corresponding to y. Consequently, an unmatched worker and firms can block if they can produce agree to a mutually beneficial outcome. A matched worker and firm pair can also block an outcome if the sum of their payoffs is lower than the total surplus they produce. The correspondence between many-to-one stable outcomes and competitive equilibria is noted in Camina (2006).

Now, we are ready to prove the desired result from which the one-to-one matching case follows trivially by allowing for only one position at each firm.

Proposition E.4 *The equilibrium assignment of positions for the games a_{ij}^f and $a_{ij}^{\tilde{f}}$ coincide. Further, if u_i^f and v_p^f are position payoffs for the game a^f, then $u_i^{\tilde{f}} = u_i^f + \left(\tilde{f}(h_i) - f(h_i)\right)$ and*

$v_p^{\tilde{f}} = v_p^f$ *are equilibrium payoffs under the surplus* $a_{ij}^{\tilde{f}}$. *Consequently the implicit tuition for each position is the same for the games* a^f *and* $a^{\tilde{f}}$.

Proof. Sotomayor (1999) shows that equilibria for a^f and $a^{\tilde{f}}$ exist and maximize the total surplus in the set of feasible assignments. Towards a contradiction, assume that $y^{\tilde{f}}$ is an equilibrium for $a^{\tilde{f}}$ but not for a^f. The feasibility constraints are identical in the two games, and so both y^f and $y^{\tilde{f}}$ are feasible for both games. Since $y^{\tilde{f}}$ (but not y^f) maximizes the total surplus under $a^{\tilde{f}}$,

$$\begin{aligned} \sum_{i,p} a_{ij_p}^{\tilde{f}} y_{ip}^{\tilde{f}} &> \sum_{i,p} a_{ij}^{\tilde{f}} y_{ip}^{f} \\ \Rightarrow \sum_{i,p} a_{ij_p}^{f} y_{ip}^{\tilde{f}} + \sum_i \sum_p \left(\tilde{f}(h_i) - f(h_i)\right) y_{ip}^{\tilde{f}} &> \sum_{i,p} a_{ij_p}^{f} y_{ip}^{f} + \sum_i \sum_p \left(\tilde{f}(h_i) - f(h_i)\right) y_{ip}^{f}. \end{aligned} \quad \text{(E.6)}$$

Since every worker-firm pair produces positive surplus and the total capacity exceeds the number of workers, there cannot be any unassigned workers in any feasible surplus maximizing allocation, i.e. $\sum_p y_{ip}^f = \sum_p y_{ip}^{\tilde{f}} = 1$ for all i. Hence, we have that $\sum_p \left(\tilde{f}(h_i) - f(h_i)\right) y_{ip}^{\tilde{f}} = \sum_p \left(\tilde{f}(h_i) - f(h_i)\right) y_{ij}^{f}$. The inequality in equation (E.6) reduces to $\sum_{i,p} a_{ij_p}^f y_{ip}^{\tilde{f}} > \sum_{i,p} a_{ij_p}^f y_{ip}^f$, a contradiction to the assumption that y^f is an equilibrium assignment for a^f. This contradiction implies that the equilibrium assignments of positions under the two games coincide.

To show that the second part of the result, consider the equilibrium payoffs for a^{f^*} where $f^*(h_i) = \max\left\{\tilde{f}(h_i), f(h_i)\right\}$ are part of an equilibrium. I show that $u_i^{f^*} = u_i^f + (f^*(h_i) - f(h_i))$ and $v_p^{f^*} = v_p^f$. The comparison of equilibrium payoffs for $\tilde{f}$ and f follows immediately from this. Note that for all i and p, $u_i^f \geq 0$ and $v_j^f \geq 0$ implies $v_j^{f^*} \geq 0$ and $u_i^{f^*} \geq 0$ since $f^*(h_i) - f(h_i) \geq 0$. It remains to show that $u_i^{f^*} + v_p^{f^*} \geq a_{ij_p}^{f^*}$ if i is assigned to position p or if i is not assigned to firm j_p. Note that for all i and p, we have that if $u_i^f + v_p^f \geq a_{ip}^f$,

$$\begin{aligned} u_i^{f^*} + v_p^{f^*} &= u_i^f + f^*(h_i) - f(h_i) + v_p^f \\ &\geq a_{ij_p}^f + f^*(h_i) - f(h_i) \\ &= a_{ij_p}^{f^*}. \end{aligned}$$

To complete the proof, I need to show that the payoffs to each position coincides under the worker-optimal stable outcome. Let u_i^f and v_p^f denote this outcome for the game a^f. Let u_i^0 and v_p^0 be the worker-optimal outcome under the function $\tilde{f}(h_i) = 0$ for all h_i. I showed earlier that the optimal assignments coincide for these two cases. I have therefore shown that $u_i^0 + f(h_i)$ and v_p^0 is stable for a^f. Towards a contradiction, assume that $u_i^f \geq u_i^0 + f(h_i)$ with strict inequality for at least one i. This implies that $u_i^f - f(h_i)$ is stable for a^0. Hence, $u_i^f - f(h_i) \geq u_i^0$ with strict inequality for at least one i, contradicting the assumption that u_i^0 and v_p^0 are part of the worker-optimal outcome. If y is the optimal assignment, this shows that $v_p^0 = \sum_i y_{ip}\left(a_{ip}^0 - u_i^0\right) = \sum_i y_{ip}\left(a_{ip}^f - u_i^f\right) = v_p^f$, proving the result. ■

E.4 Worker Optimal Equilibrium: Algorithm

The first step uses a linear program to solve for the assignment that produces the maximum total surplus. Let a_{ij} be the total surplus produced by the match of resident i with program j. This surplus is the sum of the value of the product produced by resident i at program j and the dollar value of resident i's utility for program j at a salary of 0.[5] With an abuse of notation of the letter x, let x_{ij} denote the (fraction) of resident i that is matched with program j. Sotomayor (1999) shows that the surplus maximizing (fractional) assignment is the solution to the linear program

$$\begin{aligned}&\max_{\{x_{ij}\}} \sum x_{ij} a_{ij} \\ &\text{subject to} \\ &0 \leq x_{ij} \leq 1 \\ &\sum_j x_{ij} \leq 1 \\ &\sum_i x_{ij} \leq c_j.\end{aligned} \tag{E.7}$$

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

The second step seeks to find the worker optimal salaries in any outcome with the optimal assignment $\left\{x^*_{ij}\right\}$ found in the first step. Let $\left\{y^*_{ip}\right\}$ be an associated optimal assignment to positions. An outcome $((u, v), y)$ is stable if and only if it satisfies the following linear constraints:

$$\begin{aligned}&u_i \geq 0,\ v_p \geq 0 \\ &\sum u_i + \sum v_p \leq \sum y^*_{ip} a_{ij_p} \\ &u_i + v_p = a_{ij_p} \text{ if } y^*_{ip} = 1 \\ &u_i + v_p \geq a_{ij_k} \text{ if } x^*_{ij_p} = 0.\end{aligned}$$

The first constraint is individual rationality for i and p. The second constraint is implied by the optimality of the assignment y^* as no feasible imputation may provide a larger total surplus. The third constraint asserts that the imputations supporting y^* result from loss-less transfers between a resident her matched program. The final constraints are no blocking constraints between worker i and a position at an unmatched program.

Hence, the worker optimal allocation $((u^*, v^*), y^*)$ maximizes the total worker surplus subject

[5] As mentioned in footnote 36, I assume that the equilibrium is characterized by full employment. If utilities are normalized so that an allocation is individual rationality if the resident obtains non-negative utility, then α_{ij} at the resident i's least preferred program j must exceed the negative of the dollar monetized utility resident i obtains at j at a salary of zero.

to these constraints. The solution can be obtained using a linear program since the constraints and the objective function are linear in the arguments u_i and v_p. In the counterfactual exercises, the linear programs were solved using Gurobi Optimizer (http://www.gurobi.com). Calculating the transfers implied by a solution to this problem is straightforward.

This step of the algorithm is based on the dual formulation of the one-to-one assignment problem, which has an economic interpretation given by Shapley and Shubik (1971). Sotomayor (1999) constructs the dual formulation of the many-to-one problem.

References

Agarwal, Nikhil, and William F Diamond. 2014. "Identification and Estimation in Two-sided Matching Markets." *Cowles Foundation Discussion Paper No. 1905.*

Anderson, G F. 1996. "What Does Not Explain the Variation in the Direct Costs of Graduate Medical Education." *Academic Medicine*, 71(2): 164–9.

Camina, Ester. 2006. "A generalized assignment game." *Mathematical Social Sciences*, 52(2): 152 – 161.

Gentile Jr., Thomas C, and David Buckley. 2009. "Medicare Reimbursement and Graduate Medical Education." In *Guide to Medical Education in the Teaching Hospital.* . 4th ed., , ed. J Levine. Association for Hospital Medical Education.

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

Gourieroux, Christian, and Alain Monfort. 1997. *Simulation-based Econometric Methods.* New York:Oxford University Press.

Johnson, Steven G. 2011. "The NLopt nonlinear-optimization package."

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

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

Pakes, Ariel, and David Pollard. 1989. "Simulation and the Asymptotics of Optimization Estimators." *Econometrica*, 57(5): 1027 – 57.

Roth, Alvin E., and Marilda A. Oliveira Sotomayor. 1992. *Two-Sided Matching.* Cambridge University Press.

Rowan, Thomas Harvey. 1990. "Functional Stability Analysis Of Numerical Algorithms." PhD diss.

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

Sotomayor, Marilda. 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.

Table A.1: Sample Construction

Year	03-04	04-05	05-06	06-07	07-08	08-09	09-10	10-11
Panel A: Programs								
Total number of ACGME Programs	475	462	463	460	457	453	451	451
Excluding programs in Peurto Rico	469	458	459	456	453	449	448	448
Excluding military programs	455	444	445	442	440	436	432	434
Excluding programs that do not participate in the NRMP	446	438	443	438	432	432	427	429
Excluding programs that are not offering positions	445	438	441	438	431	432	427	429
Excluding programs with no matches in the sample period	425	433	439	436	427	430	423	428
Panel B: Residents								
Total number of Residents in ACGME programs	3118	3066	3166	3148	3095	3154	3133	3268
Excluding residents matched with Peurto Rico programs	3097	3048	3154	3140	3085	3143	3126	3254
Excluding residents matched with military programs	2995	2945	3041	3026	2996	3051	3009	3160
Excluding residents matched with NRMP non-participants	2976	2925	3035	3021	2974	3040	2996	3148

Table A.2: Capacity Adjustments

Year	Number of program capacities adjusted	Average adjustment	Maximum adjusment
2003-2004	51	1.25	3
2004-2005	53	1.32	5
2005-2006	72	1.32	4
2006-2007	57	1.14	2
2007-2008	74	1.35	5
2008-2009	67	1.40	4
2009-2010	65	1.35	5
2010-2011	71	1.54	6

Notes: Capacities are adjusted upwards only. Average adjustment is reported conditional on adjustment.

Table B.1: Mis-specified Heterogeneity

		α		β		Pseudo R^2 Bias	
δ_α	δ_β	Bias	RMSE	Bias	RMSE	α	β
0	0	0.0008	0.0265	0.0042	0.0749	0.0002	0.0007
0.1	0	-0.0010	0.0258	0.0038	0.0721	0.0018	0.0006
0.5	0	-0.0294	0.0386	0.0046	0.0723	0.0405	0.0010
1	0	-0.0941	0.0973	0.0044	0.0742	0.1172	0.0008
0	0.1	-0.0010	0.0258	0.0041	0.0729	-0.0007	0.0032
0.1	0.1	-0.0044	0.0267	0.0055	0.0733	0.0001	0.0039
0.5	0.1	-0.0335	0.0418	0.0044	0.0725	0.0384	0.0034
1	0.1	-0.0975	0.1006	0.0047	0.0763	0.1154	0.0034
0	0.5	-0.0188	0.0313	0.0039	0.0733	-0.0096	0.0562
0.1	0.5	-0.0229	0.0344	0.0028	0.0750	-0.0093	0.0556
0.5	0.5	-0.0836	0.0874	0.0061	0.0771	0.0118	0.0571
1	0.5	-0.1598	0.1616	0.0076	0.0795	0.0804	0.0578
0	1	-0.0603	0.0651	-0.0014	0.0756	-0.0312	0.1645
0.1	1	-0.0648	0.0695	-0.0017	0.0756	-0.0311	0.1644
0.5	1	-0.1389	0.1412	-0.0005	0.0796	-0.0188	0.1649
1	1	-0.2537	0.2549	0.0009	0.0812	0.0243	0.1655

Table B.2: Interviewing Frictions

ρ	α		β		Number Unmatched		Pseudo R^2 Bias	
	Bias	RMSE	Bias	RMSE	Mean	s.d.	α	β
5%	-0.0003	0.0258	0.0037	0.0740	0.327	0.563	-0.0003	0.0005
10%	-0.0001	0.0258	0.0015	0.0740	0.658	0.787	-0.0002	-0.0006
20%	-0.0001	0.0256	-0.0029	0.0732	1.256	0.990	-0.0002	-0.0028
40%	-0.0003	0.0255	-0.0092	0.0726	2.301	1.268	-0.0003	-0.0059

Table B.3: Random Shuffle

κ	α		β		Pseudo R^2 Bias	
	Bias	RMSE	Bias	RMSE	α	β
5%	-0.0935	0.0974	0.0049	0.0750	-0.0491	0.0011
10%	-0.1752	0.1776	0.0050	0.0768	-0.0953	0.0010
20%	-0.3148	0.3162	0.0065	0.0849	-0.1806	0.0015
40%	-0.5311	0.5319	0.0117	0.1131	-0.3196	0.0027

Table C.1: Medicare Reimbursement Rates on Characteristics

Dependent Variable	Log Competitor Reimbursements		Log Reimbursements	
	(1)	(2)	(3)	(4)
Log Rent	-0.0057	-0.0282	-0.0004	-0.2023
	(0.0632)	(0.0737)	(0.1219)	(0.1579)
Log Wage Index	0.3924***	0.3425***	0.7497***	0.6509***
	(0.1036)	(0.1038)	(0.1977)	(0.2042)
Log Reimbursement	0.1701***	0.1538***		
	(0.0227)	(0.0221)		
Log College Share		-0.0059		0.0071
		(0.0111)		(0.0241)
Log Income		0.0025		0.0782
		(0.0648)		(0.1250)
Median Age		0.0009**		0.0013**
		(0.0003)		(0.0006)
Log Popoulation		0.0062		0.0327
		(0.0113)		(0.0258)
Violent Crime Rate		0.0017		-0.0035
		(0.0023)		(0.0070)
Property Crime Rate		-0.0024***		-0.0005
		(0.0006)		(0.0015)
Other Program Controls	Y	Y	Y	Y
Observations	3,441	3,441	3,441	3,441
R-squared	0.2335	0.2934	0.3528	0.3731

Notes: Linear regressions. Location characteristics are the median age (county), log median household income (county), log total population (MSA/county), violent crime and property crime rates (counts per 10 million) from FBI's Crime Statistics/UCR (25 mi radius weighted by 1/distance), dummies for no data in that radius and log college share (MSA/rest of state). All columns include a constant term, log # beds, log NIH Fund (Major), log NIH Fund (Minor), Log Case Mix Index, Program Type Dummies, Rural Program Dummy and dummies for programs with no NIH funding at major affiliates, for no NIH funding at minor affiliates, and a dummy for missing Medicare ID at program institutions. The Competitor Reimbursement is a weighted average of the Medicare primary care per resident amounts of institutions in the geographic area of a program other than the primary institutional affiliate of the program. Geographic area defined as in Medicare DGME payments: MSA/NECMA or Rest of State unless less than 3 other observations constitute the area, in which case the census division is used. See data appendix for description of variables and details on the construction of the reimbursement variables. A program's reimbursement rate is truncated below at $5,000 and a dummy for these 46 truncated observations is estimated as well. Standard errors clustered at the program level in parenthesis. Significance at 90% (*), 95% (**) and 99% (***) confidence.

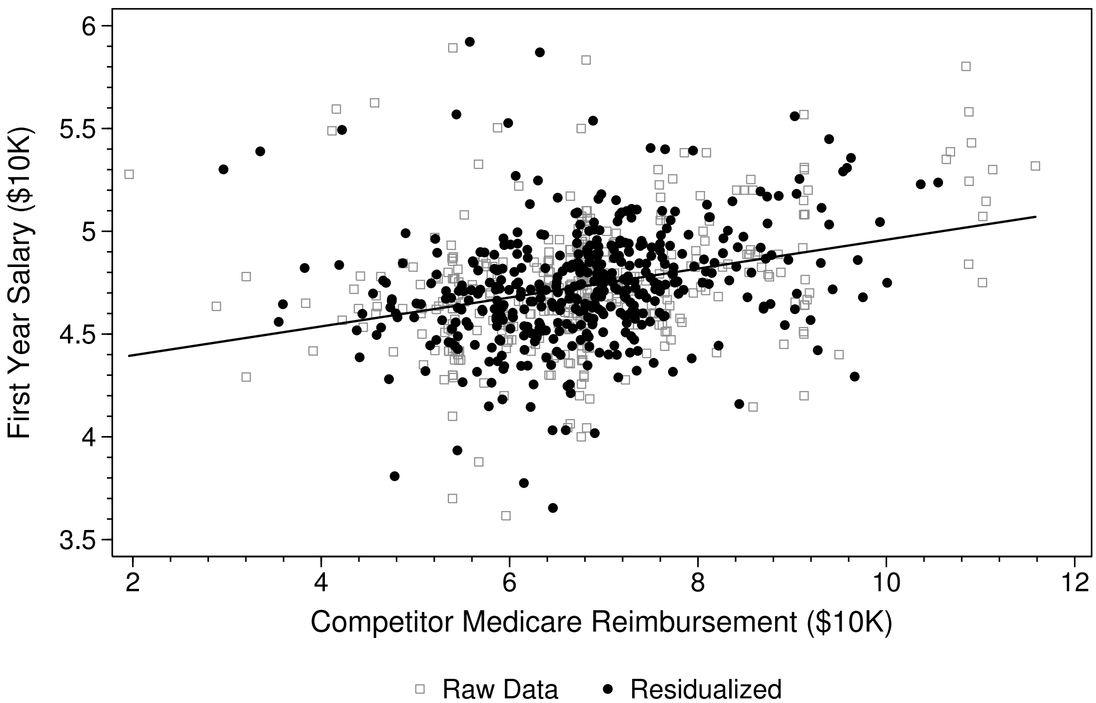

Figure C.1: Relationship Between Wages and Competitor Reimbursements

Notes: Sample restricted academic year 2010-2011. To construct the residualized scatter plot, I first regressed the X-axis and Y-axis variables on County Median Rent (Gross), Rural Program, Medicare Wage Index, Log NIH Fund. (Major), Log NIH Fund. (Minor), Log # Beds, Medicare Case-Mix Index and dummies for No NIH Fund. (Major), No NIH Fund. (Minor), missing Medicare ID. The X-axis and Y-axis residuals estimated from these regressions are scattered.

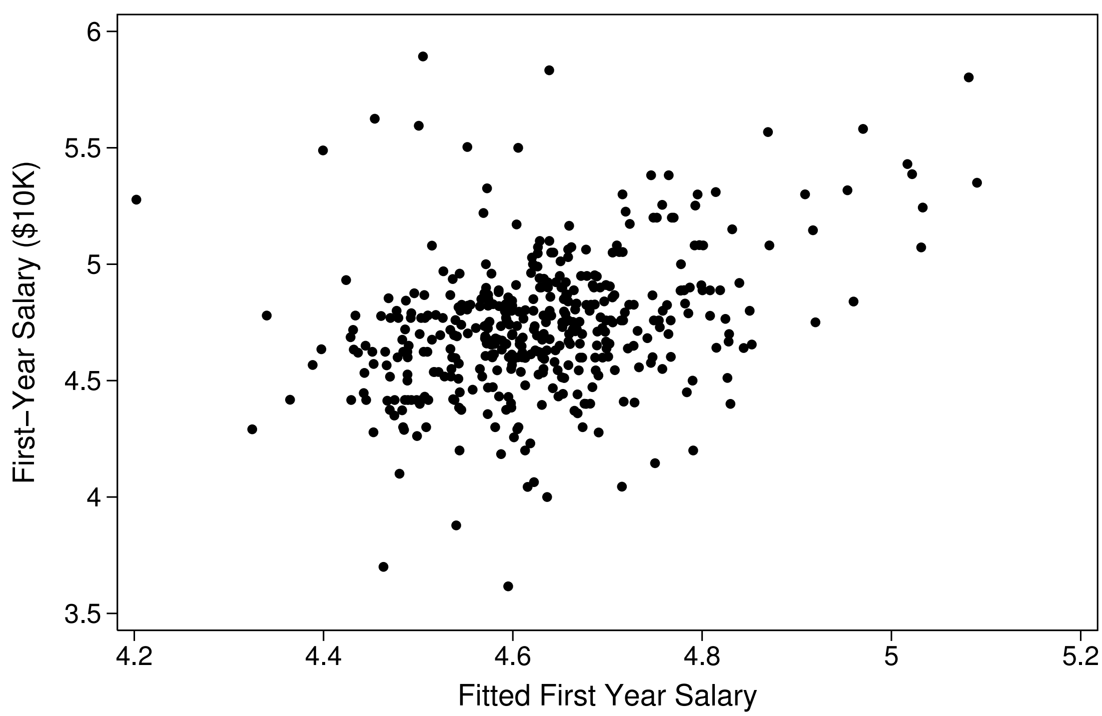

Figure C.2: Heteroskedasticity in First Stage Residuals

Notes: To construct the fitted salaries, I regressed the First Year Salary on Competitor Reimbursements, County Median Rent (Gross), Rural Program, Medicare Wage Index, Log NIH Fund (Major), Log NIH Fund (Minor), Log # Beds, Medicare Case-Mix Index and dummies for No NIH Fund (Major), No NIH Fund (Minor), missing Medicare ID. The regression was estimated on the full sample from the academic years 2002-2003 to 2010-2011. The scatter plot shows the salaries and fitted values from the academic year 2010-2011 alone. The Competitor Reimbursement is a weighted average of the Medicare primary care per resident amounts of institutions in the geographic area of a program other than the primary institutional affiliate of the program. Geographic area defined as in Medicare DGME payments: MSA/NECMA unless less than 3 other observations constitute the area, in which case the census division is used. See data appendix for description of variables and details on the construction of the reimbursement variables.

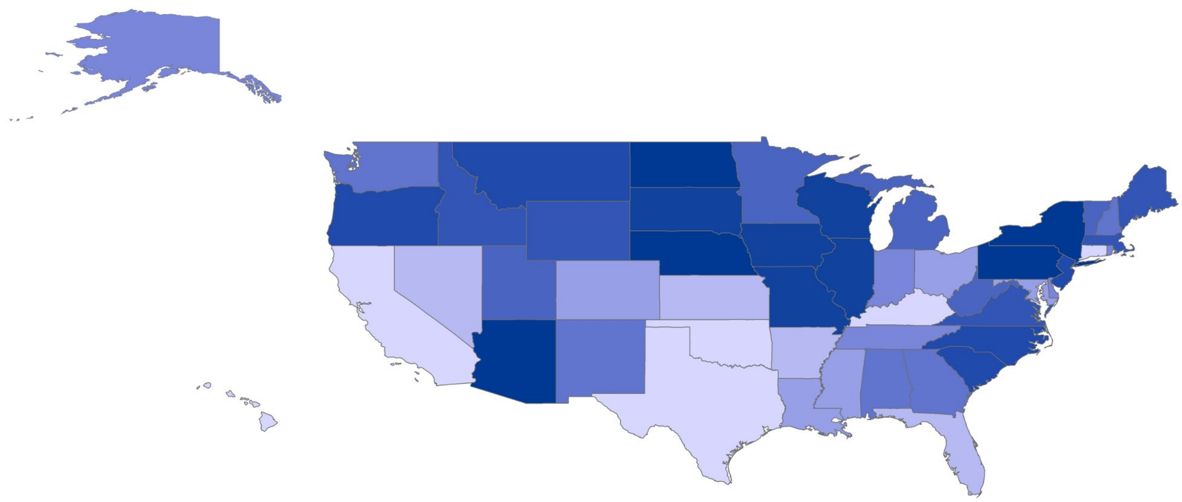

Figure C.3: Geographic Distribution of Competitor Reimbursements

Notes: Average residuals of the Competitor Medicare Reimbursements by state. Colors categorized by 10 equally sized quantiles with darker colors indicating higher values. Program sample restricted academic year 2010-2011. To construct the average residuals by state, I first regressed Competitor Medicare Reimbursements on County Median Rent (Gross), Rural Program, Medicare Wage Index, Log NIH Fund (Major), Log NIH Fund (Minor), Log # Beds, Medicare Case-Mix Index and dummies for No NIH Fund (Major), No NIH Fund (Minor), missing Medicare ID. The estimated from these regressions were averaged by the state a program is located in. The Competitor Reimbursement is a weighted average of the Medicare primary care per resident amounts of institutions in the geographic area of a program other than the primary institutional affiliate of the program. Geographic area defined as in Medicare DGME payments: MSA/NECMA unless less than 3 other observations constitute the area, in which case the census division is used. See data appendix for description of variables and details on the construction of the reimbursement variables.

Table D.1: Detailed Preference Estimates

	w/o Wage Instruments		w/ Wage Instruments	
	Geo. Het.	Full Het.	Geo. Het.	Full Het.
	(1)	(2)	(3)	(4)
Panel A: Preference for Programs				
First Year Salary ($10,000)	4.5888	2.3099	1.9531	0.4983
	(0.4500)	(0.3205)	(0.3533)	(0.3174)
Log Beds (Primary Inst)	2.6058	2.5652	2.7780	1.6392
	(0.2213)	(0.3371)	(0.2399)	(0.2656)
Log NIH Fund. (Major)	2.3046	0.0876	0.6645	-0.0474
	(0.1646)	(0.1284)	(0.0735)	(0.1350)
Log NIH Fund. (Minor)	2.2898	1.0351	1.3357	1.3589
	(0.1410)	(0.1272)	(0.1447)	(0.1461)
Medicare Case Mix Index	4.7917	4.9815	5.3517	7.9283
	(0.5733)	(0.6724)	(0.5163)	(0.9053)
Medicare Wage Index	1.9601	-5.5213	1.4322	-5.1235
	(0.5107)	(1.0418)	(0.3742)	(0.9917)
Annual Median Rent ($10,000)	-0.5741	5.9901	6.1311	7.1745
	(0.3137)	(0.8155)	(0.6117)	(0.7448)
Rural Program	2.5747	1.6925	3.3816	1.2727
	(0.3540)	(0.3457)	(0.4332)	(0.3573)
University Based Program	5.0845	3.6464	4.9082	3.6610
	(0.5451)	(0.4098)	(0.5636)	(0.4372)
Community/University Program	-1.0174	-1.1552	-1.4662	-1.7033
	(0.1645)	(0.1969)	(0.2114)	(0.2180)
Reimbursement Rate			0.2569	-0.0966
			(0.0433)	(0.0466)
Control Variable			8.7394	2.4889
			(0.7762)	(0.5335)
Rural Progam x Rural Born Resident	0.0500	0.2746	0.0455	0.2484
	(0.0113)	(0.0476)	(0.0093)	(0.0506)
Program in Medical School State	1.0563	2.2682	0.8846	2.2592
	(0.0747)	(0.1869)	(0.0555)	(0.1950)
Program in Birth State	0.6057	1.4650	0.4787	1.4643
	(0.0443)	(0.1250)	(0.0296)	(0.1269)
Sigma Log NIH Fund. (Major)		0.9814		1.1229
		(0.1833)		(0.1928)
Sigma Log Beds		4.1294		3.8453
		(0.5608)		(0.5114)
Sigma Medicare Case Mix		4.6807		3.2150
		(0.9656)		(0.9127)

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

	w/o Wage Instruments		w/ Wage Instruments	
	Geo. Het.	Full Het.	Geo. Het.	Full Het.
	(1)	(2)	(3)	(4)
Panel B: Human Capital				
Log NIH Fund (MD)	0.1269	0.1153	0.0941	0.1191
	(0.0139)	(0.0164)	(0.0131)	(0.0156)
Median MCAT (MD)	0.0666	0.0814	0.0413	0.0797
	(0.0038)	(0.0070)	(0.0030)	(0.0056)
US Born (Foreign Grad)	-0.2470	0.1503	0.2927	0.2083
	(0.0801)	(0.1021)	(0.0705)	(0.0989)
Sigma (DO)	0.7944	0.8845	0.7275	0.9321
	(0.0285)	(0.0359)	(0.0292)	(0.0370)
Sigma (Foreign)	3.0709	3.6190	2.8215	3.5549
	(0.1102)	(0.1469)	(0.1131)	(0.1411)
Medical School Type Dummies	Y	Y	Y	Y
Moments	106	106	118	118
Parameters	22	25	24	27
Objective Function	1122.78	951.31	1090.10	1032.24

Notes: See Table 7 for Panel A estimates monetized in dollar units. Indicators for zero NIH funding of major associates and for minor associates are included. In uninstrumented specifications, the variance of the vertical unobservable ξ_{jt} is normalized to 1 and in instrumented specifications, the variance of ζ_{jt} is normalized to 1. In all specifications, the variance of unobservable determinants of the human capital index of MD graduates is normalized to 1. All specifications normalize the mean utility from a program with zeros on all characteristics to 0. All specifications normalize the mean human capital index of residents with zeros for all characteristics to 0. Point estimates using 1000 simulation draws. Standard errors in parenthesis. Optimization and estimation details described in Appendix B.

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

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

Notes: Linear Regressions using 2011-2012 data. Each simulation draws a parameter from the estimated asymptotic distribution of Specification (2), and unobservables independently. The vector of coefficients is computed for each draws. The table reports the mean estimate and bootstrapped standard error of simulated estimates in parenthesis.