# Appendices

# A Estimation and Inference

## A.1 Moments

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

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

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

The three sets of moments discussed in Section 6.2 have the following mathematical expressions.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

## A.2 A Bootstrap

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

Agarwal and Diamond (2014) study asymptotic theory under this sampling process in the case of a single market $J \to \infty$. Limit theorems for the estimator is not yet complete. Monte Carlo simulations based on inference procedures for standard simulation estimators for the model with exogenous characteristics and preference heterogeneity have a decreasing root mean square error with increase in sample size. In these simulations, I used a parametric bootstrap that accounts for the dependent data structure to estimate the asymptotic variance of the moments, and a delta method to estimate the asymptotic variance of the parameter.

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

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

$$\begin{aligned}
\sqrt{J}\left(\hat{\theta}_{msm} - \theta_0\right) &\approx \left[\left(\Gamma' W \Gamma\right)^{-1} \Gamma' W\right] \sqrt{J}\left(\hat{m}\left(\hat{\theta}_{msm}\right) - m(\theta_0)\right) \\
&\xrightarrow{d} N(0, \Sigma) \\
\Sigma &= \left(\Gamma' W \Gamma\right)^{-1} \Gamma' W V^{tot} W' \Gamma \left(\Gamma' W \Gamma\right)^{-1} \\
V^{tot} &= V + \frac{1}{S} V^S \qquad (20)
\end{aligned}$$

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

To estimate $\Gamma(\theta_0)$, 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, extremely small step sizes and a low number of simulation draws can lead to inaccuracies. For this step, I use 10,000 simulation draws and a step size of $10^{-3}$. The simulation variance is estimated by calculating the variance in 10,000 evaluations of $\hat{m}^S\left(\hat{\theta}_{msm}\right)$, each with a single simulation draw and using the observed sample of resident and program characteristics. Since these two calculations keep the set of observed residents and programs constant, these two quantities can be calculated independently in each of the markets.

As noted, the sampling variance in $\hat{m}(\theta)$ 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 $\{z_{jt}, r_{jt}, q_{jt}\}_{j=1}^{J_t}$ with replacement. Denote this sample with $\left\{z_{jt}^b, r_{jt}^b, q_{jt}^b\right\}_{j=1}^{J_t}$

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

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

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

Essentially, the bootstrap mimics the data generating process to sample a new set of agents from the population distribution to form an economy. It replaces the set of observed characteristics of the residents and programs with the empirical distribution observed in the data. Given this economy, it computes $\hat{\nu}_{jt}$ and the moments at a pairwise stable match at $\hat{\theta}$. The covariance of the moments across bootstrap iterations is the estimate of $\hat{V}$. The uncertainty due to simulation error $\hat{V}^S$ is approximated by drawing just the unobserved characteristics.[47] In a large economy, consistency of each of these quantities implies the consistency of the estimate

$$\hat{\Sigma} = \left(\hat{\Gamma}'W\hat{\Gamma}\right)^{-1}\hat{\Gamma}'W\left(\hat{V} + \frac{1}{S}\hat{V}^S\right)W'\hat{\Gamma}\left(\hat{\Gamma}'W\hat{\Gamma}\right)^{-1}. \tag{21}$$

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

### A.2.1 Weight Matrix

It is well known that the choice of weight matrix can affect efficiency. This choice is particularly important when the number of moments is much larger than the number of parameters. A common method uses a first stage consistent estimate of $\theta_0$ to obtain variance estimates $\hat{V}$ and $\hat{V}^S$ to compute the optimal weight matrix $\hat{W} = \left(\hat{V} + \frac{1}{S}\hat{V}^S\right)^{-1}$ that can be used in the second stage. One may implement the first step of obtaining a consistent estimate of $\theta_0$ using any positive definite matrix $W$, with the identity matrix as the most commonly used first-step weight matrix. In this application, a two-step procedure is computationally prohibitive. In Monte Carlo simulations with this dataset, I found that using the identity matrix was often inaccurate and left us with a poor estimate of $\theta_0$. Instead, a weight matrix $\tilde{W}$ calculated using the following bootstrap procedure seemed to approximate the optimal weights fairly well. For each market $t$, with replacement, randomly sample $J_t$ programs and the residents matched with them. Treat the observed matches as the matches in the bootstrap sample as well.[48] Compute moments $\left\{\tilde{m}^b\right\}_{b=1}^B$ from the sample and compute the variance $\tilde{V}$ and set $\tilde{W} = \tilde{V}^{-1}$. While this weight matrix need not converge to the optimal weight matrix, the only theoretical loss is in the efficiency of the estimator. This weight matrix also turns out to be close to one that would be calculated as $\hat{W} = \left(\tilde{V}\left(\hat{\theta}_{msm}\right) + \frac{1}{S}\hat{V}^S\left(\hat{\theta}_{msm}\right)\right)^{-1}$ where $\hat{\theta}_{msm}$ is the estimate of $\theta_0$ using $\hat{W}^{sub}$ as the weight matrix, and $\tilde{V}\left(\hat{\theta}_{msm}\right)$ and $\hat{V}^S\left(\hat{\theta}_{msm}\right)$ are the sample and simulation variance that are estimated as described earlier.

## A.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 $\hat{\theta}_{msm}$.

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

As with the vast majority of optimizers working with non-convex problems, there is no guarantee that the genetic algorithm finds the global optimum. I conducted three initial genetic algorithm runs to with separately seeded populations of size 40, cross-over fraction

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

of 0.75, one elite child, an adaptive mutation scale of 4 and shrinkage of 0.25. These extensive runs were used to generate starting values for the local searches.

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

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

# B Parameter Estimates

Table B.1 presents point estimates of the models discussed in Section 7 and three additional models. Two of the additional models do not allow for heterogeneity in preferences. The final additional model is a version of specification (1) in Table 8 that uses the instrument.

Panel A presents parameter estimates for the distribution of residents' preferences and Panel B presents estimates for the human capital index. As mentioned in the text, these point estimates are not directly interpretable in economically meaningful terms. Table 8 translates a subset of coefficients from Panel A into monetized values by dividing a given coefficient by the coefficient on salaries, and scaling them into dollar equivalents for a one standard deviation change.

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

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

Comparing estimates from specifications (1) and (2), we see changes in the estimated coefficient on NIH funding of major affiliates, salaries and the medicare wage index, and rent. Note that the change in coefficient on rent does not appear to have economically meaningful impact on the willingness to pay for programs located in high rent areas as compared to programs in low rent areas. Table 9 shows that specifications (1) and (2) yield similar quantities on this front. A reason for this is that medicare wage index and rents are highly correlated with each other. We also see that the relative magnitude on coefficients on rural birth interacted with rural program, program location in birth state and program location in medical school state have similar relative magnitudes although large in overall magnitude in specification (1). I attribute this difference to additional unobserved heterogeneity in specification (1), due to which similar geographic sorting needs to be explained with higher preference for these characteristics.

# C Wage Competition

## C.1 Expressions for Competitive Outcomes

I first characterize the competitive equilibria of the model. The expression in equation (17) follows as a corollary. For clarity, I refer to the quality of program 1 as $q_1$ although I normalize it to 0 in the model presented in the text.

**Proposition 3** *The wage $w_k$ 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 wages 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,

$$\begin{aligned} f\left(h_1, q_1\right) - w_1 &\geq f\left(h_k, q_1\right) - w_k + a\left(q_1 - q_k\right) \\ \Rightarrow w_k &\geq w_1 + f\left(h_k, q_1\right) - f\left(h_1, q_1\right) + a\left(q_1 - q_k\right) \\ &\geq -aq_k \end{aligned} \tag{22}$$

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

$$\begin{aligned} f\left(h_k, q_k\right) - w_k &\geq f\left(h_1, q_k\right) - w_1 + a\left(q_k - q_1\right) \\ \Rightarrow w_k &\leq f\left(h_k, q_k\right) - f\left(h_1, q_k\right) + w_1 - a\left(q_k - q_1\right) \\ &\leq f\left(h_k, q_k\right) - f\left(h_1, q_1\right) + w_1 - a\left(q_k - q_1\right) \\ &\leq f\left(h_k, q_k\right) \end{aligned} \tag{23}$$

where the last two inequalities follow since $w_1 \leq f\left(h_1, q_1\right)$ from $IR\left(1\right)$ and $-a\left(q_k - q_1\right) \leq 0$. Equations (22) and (23) imply $IR\left(k\right)$.

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

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

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

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

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

$$\begin{aligned} f(h_k, q_k) - w_k &\geq \sum_{i=m+1}^{k} [f(h_{m-1}, q_l) - f(h_{m-1}, q_{i-1})] + f(h_{m-1}, q_m) + a(q_k - q_{m-1}) - w_{m-1} \\ &= f(h_{m-1}, q_k) + a(q_k - q_{m-1}) - w_{m-1}. \end{aligned} \tag{24}$$

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 4** *The worker optimal competitive equilibrium wages are given by*

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

*and the firm optimal competitive equilibrium wages are given by*

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

## C.2 Proposition 1

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

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

**Theorem 5** *In the worker-optimal (firm-optimal) competitive equilibria, each worker's wage under $f \in \mathcal{F}(y)$ is bounded above by her wage under $\bar{f}_y$ and below by her wage under $\underline{f_y}$.*

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

**Proof.** I only derive the bounds for the worker optimal equilibrium since the calculation for the firm optimal equilibrium is analogous. From the expressions in corollary 4,

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

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

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

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

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

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

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

■

Proposition 1 follows as a corollary.

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

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

■

## C.3 Worker Optimal Equilibrium: Algorithm

The first step uses a linear program to solve for the assignment that produces the maximum total surplus. Let $a_{ij}$ be the total surplus produced by the match of resident $i$ with program $j$. This surplus is the sum of the value of the product produced by resident $i$ at program $j$ and the dollar value of resident $i$'s utility for program $j$ at a wage of 0.[49] With an abuse of notation of the letter $x$, let $x_{ij}$ denote the (fraction) of resident $i$ that is matched with program $j$. Sotomayor (1999) shows that the surplus maximizing (fractional) matching is

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

the solution to the linear program

$$\begin{aligned}
&\max_{\{x_{ij}\}} \sum x_{ij} a_{ij} \\
&\text{subject to} \\
&0 \leq x_{ij} \leq 1 \\
&\sum_j x_{ij} \leq 1 \\
&\sum_i x_{ij} \leq c_j.
\end{aligned} \tag{25}$$

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

The second step seeks to find the worker optimal wages supporting this assignment. The algorithm is based on the dual formulation of the one-to-one assignment problem, which has an economic interpretation given by Shapley and Shubik (1971). Assume for now that $c_j = 1$ for all $j$. If $u_i$ is the utility imputation for resident $i$ and $v_j$ is the imputation for program $j$, then a core allocation ensures that for all $i$, $j$ $u_i + v_j \geq a_{ij}$. This inequality holds for a core allocation if $i$ and $j$ are matched since utility in fully transferable, and if $i$ and $j$ are not matched since otherwise $i$ and $j$ would block the allocation.[50] A particular element in the core can be found by solving the problem

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

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

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

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

and $y_{ik'} = 0$ for all positions $k'$ at program $j_k$. In other words, an allocation is blocked only by a resident and position pair in which the position is at a program other than the resident's assignment.

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

$$\begin{aligned}
&\max_{\{u_i\},\{v_k\}} \sum u_i \\
&\text{subject to} \\
&u_i \geq 0,\ v_k \geq 0 \\
&\sum u_i + \sum v_k \leq \sum x^*_{ij} a_{ij} \\
&u_i + v_k = a_{ij_k} \text{ if } y^*_{ik} = 1 \\
&u_i + v_k \geq a_{ij_k} \text{ if } x^*_{ij_k} = 0.
\end{aligned} \tag{26}$$

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

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

## C.4 Implicit Tuition

I prove a more general result for many-to-one assignment games that subsumes Proposition 2. To do this, I first need to introduce some notation. A many to one assignment game between workers $i \in \{1, \ldots, N\}$ and firms $j \in \{1, \ldots, J\}$. The capacity of firm $j$ is $c_j$. I focus on the case when $\sum_j c_j \geq N$. Worker $i$, with human capital $h_i$, produces $f(h_i) \geq 0$ at firm $j$, independently of the other workers at the firm. An empty slot produces 0. The utility worker $i$ receives from working at firm $j$ at a wage of $w$ is $u_{ij} + w$. Since the wage transfer is lossless, the total surplus produced by the pair $i,\ j$ under the production function $f$ is $a^f_{ij} = u_{ij} + f(h_i)$. I assume that each $u_{ij} \geq 0$.

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

In the many-to-one case, we refer to an **assignment of positions** $\{y_{i,p}\}_{i,p}$ where $p \in$

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

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

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

**Proposition 6** *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 $u_i^{\tilde{f}} = u_i^f + \left(\tilde{f}\left(h_i\right) - f\left(h_i\right)\right)$ and $v_p^{\tilde{f}} = v_p^f$ are equilibrium payoffs under the surplus $a_{ij}^{\tilde{f}}$. Consequently the implicit tuition for each position is the same for the games $a^f$ and $a^{\tilde{f}}$.*

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

$$\begin{aligned} \sum_{i,p} a_{ij_p}^{\tilde{f}} y_{ip}^{\tilde{f}} &> \sum_{i,p} a_{ij}^{\tilde{f}} y_{ip}^{f} \\ \Rightarrow \sum_{i,p} a_{ij_p}^{f} y_{ip}^{\tilde{f}} + \sum_i \sum_p \left(\tilde{f}\left(h_i\right) - f\left(h_i\right)\right) y_{ip}^{\tilde{f}} &> \sum_{i,p} a_{ij_p}^{f} y_{ip}^{f} + \sum_i \sum_p \left(\tilde{f}\left(h_i\right) - f\left(h_i\right)\right) y_{ip}^{f} \end{aligned} \quad (27)$$

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 $\sum_p \left(\tilde{f}\left(h_i\right) - f\left(h_i\right)\right) y_{ip}^{\tilde{f}} =$

$\sum_i \left(\tilde{f}(h_i) - f(h_i)\right) y_{ij}^f$. The inequality in equation (27) reduces to $\sum_{i,p} a_{ij_p}^f y_{ip}^{\tilde{f}} > \sum_{i,p} a_{ij_p}^f y_{ip}^f$, a contradiction to the assumption that $y^f$ is an equilibrium assignment for $y^f$. This contradiction implies that the equilibrium assignments of positions under the two games coincide.

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

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

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

# D Rural Hospitals

## D.1 Suggestive Evidence on Preference Heterogeneity for Rural Doctors

If preferences for resident traits other than a single human capital index were important, one expects that two residents at the same program have dissimilar academic qualifications if they differ on these dimensions. More concretely, one may expect that at rural programs, a rural born resident is academically less qualified than her peer born in an urban location. This may happen because a rural program prefers a rural born resident to an equally qualified urban born resident. To assess whether rural born residents in rural hospitals are more qualified than their urban colleagues, I estimated the regression

$$x_i = \delta \ rural_i + program\_fe_{\mu(i)} + e_i,$$

where $x_i$ is a measure of medical school quality for resident $i$, $rural_i$ is a dummy for a rural born resident and $program\_fe_{\mu(i)}$ is a fixed effect for program $\mu(i)$, resident $i$'s match.

The results presented in Table D.3 suggests that this may not be of primary importance. Columns (1) and (6) show that rural born residents matched with rural hospitals hail from medical schools that have, on average, only about 0.06 log points less NIH research funding that their peers born in urban areas and are about one percentage point more likely to have an MD degree. Note that the standard deviation in log NIH funding is 1.23. Neither estimate is statistically significant. Although not presented here, the conclusion is robust to using median MCAT score as an indicator of a resident's quality in place of research funding or medical school ranks. If program-year fixed effects are included in place of program-fixed effects, the estimates are more imprecise and the hypothesis that the medical school qualities of the rural born residents at rural hospitals is same as their urban born peers still cannot be rejected. Columns (3) and (8) of the table show this observation is despite the fact that the average rural born residents hails from an observably different medical school than their urban counterparts.

As a validation exercise, I ran similar regressions using gender in place of rural birth. Since accreditation guidelines prohibit programs from discriminating on the basis of sex,[51] one may reasonably expect that there is no gender based discrimination by residency programs. Columns (5) and (6) show that although the average female resident hails from medical schools that is better funded than male residents in their cohort, their medical school quality is no different from their male colleagues in their residency program.

While these results are reassuring, they are not definitive on the lack of preference heterogeneity. The somewhat large standard errors and the fact that these observables are proxies for resident quality are the primary reasons for this reserved interpretation. Nonetheless, they suggest that estimates may not suffer from large biases.

[51] The institutional requirements from the Acceditation Council for Graduate Medical Education (ACGME) states that "ACGME-accredited programs must not discriminate with regard to sex, race, age, religion, color, national origin, disability, or any other applicable legally protected status."

## D.2 General and Partial Equilibrium Effects of Financial Incentives

I consider a partial equilibrium alternative to simulations presented in Section 9.1 that may be analytically inexpensive but could, in some situations, perform fairly well. Suppose a policymaker could survey rural residency program directors to determine the impact of incentives for rural training on the residents that choose to train there. For instance, a survey such as the National GME census used in this study could be also solicit a program director's judgement of the number and quality of residents that would match to the program if it unilaterally raised its salary. The responses could be used to predict the impact of the financial incentives studied earlier by simply aggregating the number of positions filled and resident types in rural areas. Such a calculation ignores the influence of a resident who is on the margin between two rural programs and an urban program on the final results. By ignoring the fact that salaries at all rural programs would be increased simultaneously, the calculation acts as if program directors at both rural programs believe that this resident is matched to their program.

The hypothetical benchmark can be simulated using the estimated model by aggregating predicted changes in the matches from the unilateral salary increases at rural hospitals. Panel A of Table D.4 compares results for \$5,000, \$10,000 and \$20,000 increase in salary to rural programs. Comparing the results with those in Panel A, it appears that this simple partial equilibrium analysis would do fairly well at predicting the overall impact of subsidies to rural programs. The impact on resident quality and numbers are only slightly overstated. This observation is because at the estimated parameters, most residents are indifferent between a rural hospital and an urban hospital rather than two rural hospitals, and the number of rural positions is only about a tenth of all positions in the market. This fact is reflected in the distribution graphed in Figure 2.

Panel B of Table D.4, compares outcomes for incentives for training in rural programs as well as medically underserved states. The ACA redistributes previous allocated funding to urban programs but currently unused to residency training to (i) rural programs, (ii) states in the bottom quartile of the physician to population ratio and (iii) states in the top 10 in numbers of people living in a Health Physician Shortage Area. I label these states[52] as medically underserved states and compare the partial and general equilibrium impacts of financial incentives. We see that for a \$5,000 incentive, the partial equilibrium analysis predicts an 11% larger impact of subsidies. Notice that for larger subsidies, the difference between the partial and general equilibrium predictions in the change in the number of matches is smaller. For a larger subsidy, the partial equilibrium analysis overstates the change in quality of residents matched at programs in medically underserved states.

Qualitatively similar, but quantitatively larger answers were obtained from a simulation exercise in which I randomly subsidized one-quarter of the residency programs. Panel C presents these results. These simulation experiment shows that the model is capable of

[52] CMS identified Montana, Idaho, Alaska, Wyoming, Nevada, South Dakota, North Dakota, Mississippi, Florida, Peurto Rico, Indiana, Arizona and Georgia as in the bottom quartile of physicians to population ratio. Lousiana, Mississippi, Peurto Rico, New Mexico, South Dakota, District of Columbia, Montana, North Dakota, Wyoming and Alabama are in the top 10 in numbers of people living in primary care HPSAs. Peurto Rico is exlcuded from this analysis.

capturing potentially important general equilibrium effects of policy interventions. The size of these effects depends on the primitive preferences in the market structure as well as the scope of the intervention.

# E Data Construction

## E.1 National GME Census

The American Medical Association (AMA) and the Association of American Medical Colleges (AAMC) jointly conduct an annual National Graduate Medical Education Census (GME Track) of all residency programs accredited by the Accreditation Council for Graduate Medical Education (ACGME). There are two main components of the census: the program survey and the sponsoring institution survey. The program survey, which is completed by the program directors, also gathers information about the residents training at the programs. Fields from the surveys are used to update FRIEDA Online, a publicly accessible database and the AMA physician masterfile. Since 2000, the GME Track has been pre-augmented with data from the Electronic Residency Application Service (ERAS) and the National Residency Matching Program (NRMP).[53] The AMA provided records from the National GME census on all family medicine residency training programs in the Unites States between 2003-2004 and 2010-2011. The 2011-2012 data was provided after the initial empirical analysis was completed.

The data files and identifiers are structured as follows:

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

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

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

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

### E.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.

[53] 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.

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 E.6.

### E.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 programs occur. The AMA data identifies the relationship between programs and sponsoring institutions and hospitals in two ways. The program files records list each program's primary site. The program-institution bridge file records the sponsoring institution, (a second) primary clinical site and other affiliated institutions.

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

In a large number of cases, the 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. In any given year, no more than 43 programs were affected in any given year.

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

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

I use these relationships to define two types of affiliations for programs to medical schools, major and minor. A program has a major affiliation to a medical school if the primary or sponsoring institution has a major affiliation with a medical school. All other relationships are regarded as minor relationships. The relationships between programs and medical schools are imputed for all years between the first and last year of a major (likewise minor) relationship. I used all relationships since 1996 for this imputation and for 2010-2011, I used the relationships in 2009-2010 as well. For the unselected sample of family medicine programs between 2002-2003 and 2010-2011, I imputed relationships for 144 out of 2797 major affiliations and 702 out of 3337 minor affiliations. The mean NIH funding across all major and minor affiliations are used as the variables for this merge.

## E.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.

### E.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 include funding for projects designated to Schools of Medicine, Schools of Medicine and Dentistry, and Overall Medical as these categories were the major categories at which the recipient was affiliated with an MD medical school. I wished to include funding only for extramural and cooperative research activities, and training and fellowship programs funded by the NIH in a medical school. So, I dropped activity codes beginning with G, C, H as these were designated for construction, resource development and community service. Further, I dropped activity codes beginning with N and Z since those data are available only after 2007.

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

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

### E.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.

## E.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 G.

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

### E.3.1 Medicare Wage Index and Case Mix Index

The Center for Medicare Services calculated a Wage Index and Case Mix Index for each provider.[54] I merged the CMS data with primary institution. In a small number of instance, the primary institution did not have a match with Medicare data. In these cases, I calculated the average of the variable for all affiliates with Medicare data. In a total of 63 out of 3441 cases, the case mix index was not available even for affiliates. Here, in the structural estimates, I used an imputed value from a linear regression on all other characteristics included in the demand system. Finally, missing values of the wage index were imputed using the geographic definitions Medicare uses to calculate the wage index.

[54] 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

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

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

## E.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.

## E.6 Other Data

### E.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.

### E.6.2 Rent

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

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

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

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

## E.7 Miscellaneous Issues

1. For the preference estimates, imputation of salaries for missing data was done for 23 observations out of 3441 using a linear regression on the other characteristics included in the model.

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

   I find instances when the number of residents in first year positions exceeds this capacity measure. In these cases, I take the maximum of the number matched to the program and the lagged response to the first year enrollment as the program's capacity. In more than 75% of the cases, the number matched did not exceed the reported number of positions by more than one. Table E.7 summarizes the number of observations affected by this change and the mean size of the change. One reason for the discrepancy may be residents that repeated their first year training or deferred enrollment.

## F The Distribution of Physician Starting Salaries

The experience adjustment uses the following Mincerian wage regression to capture the impact on physician productivity:[55]

$$\ln y_i = \rho_0 + \rho_1 t_i + \rho_2 t_i^2 + \rho_c c_i + e_i. \tag{28}$$

Here, $y_i$ is the earnings of physician $i$, $t_i$ is the experience of physician $i$, $c_i$ is a vector of controls and $e_i$ is mean zero error. The functional form is motivated by a multiplicative return to human capital, which increases with job experience up to a maximum before depreciating.

I use records from the restricted-use file of family practice physicians from the Health Physician Tracking Survey of 2008 to estimate $\rho$. The survey collects data on the income category of physicians in the United States, with medical specialty, years practicing medicine and a variety of other fields related to their medical practice. The survey asks for the income earned by the physician in 2006 from medically related activities, excluding returns on investments in stocks or assets in their practice. The income field is coded into groups \$50,000, with the lowest category for physicians with an income under \$100,000 and the highest category for physicians with an income of \$300,000 or more. I use an interval regression in which $e_i \sim N(0, \sigma_e)$ to estimate $(\rho,\ \sigma_e)$.

Table F.8 presents summaries from the subpopulation of physicians under the age of 60 in 2006, the year of the income data in the survey. The vast majority of family physicians are salaried and earn \$200,000 or less. Table F.9 presents maximum likelihood estimates from the interval regression model. The point estimates evidence for concavity in returns to experience and a gender-pay gap that is well-documented in the empirical literature. A comparison of estimates in columns (2) and (3) also suggest some heteroskedasticity in the distribution of pay across experience levels. Column (4) estimates a quadratic functional form for this heteroskedasticity and finds a concave relationship, with a higher cross-sectional variation in earnings for physicians in the middle of their career than for physicians early or late in their career.

[55] See motivating theoretical model in Ben-Porath (1967), some early empirical work in Mincer (1974). Thomas Lemieux (2006) and Heckman, Lochner, and Todd (2003) survey the literature on mincer regressions.

# G Medicare Reimbursement Rates and Instrument Details

## G.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 grandfarthered 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.

## G.2 The Instrument: Competitor Reimbursement Rates

To construct competitor reimbursements, I first extract the records from the fields "Updated per resident amount for OB/GYN and primary care" and "Number of FTE residents for OB/GYN and primary care" on lines 2 and 1 respectively in form CMS-2552-96, Worksheet

E-3, Part-IV for the cost reporting period beginning October 1, 1996 and before September 30,1997. As per the instructions for this form (3633.4), this is the latest period for which the response to the field was required by the hospitals. Indeed, I found only five observations for this field in the cost reporting period ending October 1, 1998 and no observations in the next period. The per resident amount variable is recorded in cents, and so is first converted into dollars. Both fields were winsorized at the bottom at top 1 percent since the range of values were extreme. Barring the effects of winsorizing the data, the distribution of the per resident amount variable is similar to Figure G.2 taken from Newhouse and Wilensky (2001). While some institutions have per resident amounts less than $40,000, others are reimbursed at rates higher than $200,000.

The Competitor Reimbursement variable for an institution is constructed in order to mimic the per resident amount calculation done by Medicare for new sponsors. As given in equation (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 G.4 depicts the state-averaged variation in the instrument that is not explained by the controls included in the preference estimates and a program's own reimbursement rate. A degree of spatial correlation within a census division is noticeable due to the definition of the geographical units used. Table G.10 presents regressions of the instrumental variable on characteristics included in the preference estimation, as well as location characteristics such as median age, median household income, crime rates, total population and college share. These location characteristics, together with program characteristics explain only 27% of the variation in the instrument. Strictly speaking a test for exogeneity with respect to the additional location characteristics would be rejected at the 1% level. However, the location characteristics together explain only about 6% of the variation not explained by the other controls that are included in the preferences estimates. Columns (4-6) show that characteristics of the program itself explain about 35% of the variation in its reimbursement rates and the addition of location characteristics is not important. These findings are consistent with Anderson (1996), which argues against this reimbursement schemes on the basis that other cost predictors do not correlate very strongly with per resident amounts. Strictly speaking, these findings do not fully support strict exogeneity of the instrument.

Table B.1: Detailed Preference Estimates

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

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

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

Notes: See Table 8 for Panel A estimates monetized in dollar units. Indicator for zero NIH funding of major associates and for minor associates. In uninstrumented specifications, the variance of the vertical unobservable $\xi_{jt}$ is normalized to 1 and in instrumented specifications, the variance of $\zeta_{jt}$ is normalized to 1. In all specifications, the variance of unobservable determinants of the human capital index of MD graduates is normalized to 1. All specifications normalize the mean utility from a program with zeros on all characteristics to 0. All specifications normalize the mean human capital index of residents with zeros for all characteristics to 0. Point estimates using 1000 simulation draws. Standard errors in parenthesis. Optimization and estimation details described in an appendix.

Table B.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 (1), and unobservables independently. The vector of coefficients is computed for each draws. The table reports the mean estimate and bootstrapped standard error of simulated estimates in parenthesis.

Table D.3: Preferences for Rural Born Doctors

| | Log NIH Funding (MD) | | | | | Allopathic/MD Degree | | |
|---|---|---|---|---|---|---|---|---|
| | Rural Pgms. | Urban Pgms. | All | All | All | Rural Pgms. | Urban Pgms. | All |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Rural Born Resident | -0.0582 | 0.0015 | -0.1284*** | | | 0.0122 | 0.0176 | 0.0263** |
| | (0.0811) | (0.0364) | (0.0339) | | | (0.0324) | (0.0119) | (0.0107) |
| Female | | | | -0.0153 | 0.0681*** | | | |
| | | | | (0.0234) | (0.0255) | | | |
| Program Fixed Effect | Y | Y | | Y | | Y | Y | |
| Observations | 750 | 7,885 | 8,635 | 9,599 | 9,599 | 1,200 | 11,260 | 12,460 |
| R-squared | 0.2916 | 0.2461 | 0.0017 | 0.2535 | 0.0007 | 0.2568 | 0.2662 | 0.0005 |

Notes: Linear regression of resident's graduating school characteristic on other resident characteristics. Column header Rural (Urban, All) indicates regressions using residents matched to rural (urban, all) programs. Samples pooled from the academic years 2003-2004 to 2010-2011. Columns (1-5) restrict to residents graduating from medical schools with non-zero average annual NIH funding. Columns (1-3) and (6-8) restrict to the subset of residents with reliable city of birth information and were born in the United States. Standard errors clustered at the program level in parenthesis. Significance at 90% (*), 95% (**) and 99% (***) confidence.

Table D.4: General vs. Partial Equilibrium Effects of Price Incentives

| | Full Heterogeneity w/o Wage Instruments | | |
|---|---|---|---|
| Subsidy Size | \$5,000 | \$10,000 | \$20,000 |
| | (1) | (2) | (3) |
| Panel A: Rural Programs | | | |
| Total Capacity | — | 334 | — |
| Observed # Matches | — | 310 | — |
| Baseline Simulated Matches | — | 313.33 | — |
| Baseline Prob Rural Match > Urban Match | — | 52.76% | — |
| | | | |
| $\Delta$ # Matches (General Equilibrium) | 10.23 | 17.3 | 20.63 |
| $\Delta$ Prob Rural Match > Urban Match (GE) | 9.38% | 17.70% | 31.28% |
| $\Delta$ # Matches (Partial Equilibrium) | 10.31 | 17.59 | 20.63 |
| $\Delta$ Prob Rural Match > Urban Match (PE) | 10.22% | 19.56% | 34.22% |
| Panel B: Medically Underserved States and Rural Programs (MUA) | | | |
| Total Capacity | — | 751 | — |
| Observed # Matches | — | 720 | — |
| Baseline Simulated Matches | — | 721.79 | — |
| Baseline Prob MUA Match > Other Matches | — | 53.53% | — |
| | | | |
| $\Delta$ # Matches (General Equilibrium) | 14.72 | 24.7 | 29.17 |
| $\Delta$ Prob MUA Match > Other Matches (GE) | 8.73% | 16.82% | 29.93% |
| $\Delta$ # Matches (Partial Equilibrium) | 16.46 | 25.88 | 29.17 |
| $\Delta$ Prob MUA Match > Other Matches (PE) | 9.31% | 18.25% | 32.70% |
| Panel C: 1 in 4 Randomly Chosen Programs | | | |
| $\Delta$ # Matches (General Equilibrium) | 21.54 | 32.23 | 38.74 |
| $\Delta$ # Matches (Partial Equilibrium) | 25.45 | 34.04 | 39.05 |
| Prob PE Match > GE Match | 52.59% | 56.43% | 67.58% |

Notes: Medically underserved states are in the bottom quartile of physician to population ratios or in the top 10 in total area designated as a Health Physician Shortage Area (HPSA). All simulations use 2010 - 2011 sample with 3,148 residents and 3,297 total number of positions. Baseline and counterfactual simulations using 100 draws of structural unobservables. Inter-quartile range in parenthesis. Prob. X > Y is the Wilcoxian statistic: probability that the human capital of the population X is drawn from is greater than that of the population that Y is drawn from.

Table D.5: Recruitment Into Rural Practice

| | Urban Born Resident | | Rural Born Resident | |
|---|---|---|---|---|
| | Urban Program | Rural Program | Urban Program | Rural Program |
| Percent Practicing in a Rural County | 19.52% | 50.45% | 46.35% | 79.19% |

Notes: Means of location outcomes for US born residents entering a non-academic practice and with good data on birth city and practice city. Post-graduation plans from graduating resident survey administered to residency program directors in the National GME Census Track. The headers Urban (Rural) Program indicates whether a resident graduated from an urban (rural) program. Results from 5878 resident observations and 2027 program-year observations.

Table E.6: Sample Construction

| Year | 2003-2004 | 2004-2005 | 2005-2006 | 2006-2007 | 2007-2008 | 2008-2009 | 2009-2010 | 20 |
|---|---|---|---|---|---|---|---|---|
| *Panel A: Programs* | | | | | | | | |
| Total number of ACGME Programs | 475 | 462 | 463 | 460 | 457 | 453 | 451 | |
| Excluding programs in Peurto Rico | 469 | 458 | 459 | 456 | 453 | 449 | 448 | |
| Excluding military programs | 455 | 444 | 445 | 442 | 440 | 436 | 432 | |
| Excluding programs that do not participate in the NRMP | 446 | 438 | 443 | 438 | 432 | 432 | 427 | |
| Excluding programs that are not offering positions | 445 | 438 | 441 | 438 | 431 | 432 | 427 | |
| Excluding programs with no matches in the sample period | 425 | 433 | 439 | 436 | 427 | 430 | 423 | |
| | | | | | | | | |
| *Panel B: Residents* | | | | | | | | |
| Total number of Residents in ACGME programs | 3118 | 3066 | 3166 | 3148 | 3095 | 3154 | 3133 | |
| Excluding residents matched with Peurto Rico programs | 3097 | 3048 | 3154 | 3140 | 3085 | 3143 | 3126 | |
| Excluding residents matched with military programs | 2995 | 2945 | 3041 | 3026 | 2996 | 3051 | 3009 | |
| Excluding residents matched with NRMP non-participants | 2976 | 2925 | 3035 | 3021 | 2974 | 3040 | 2996 | |

Table E.7: Capacity Adjustments

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

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

Table F.8: Characteristics of Family Medicine Doctors in the US

| | Mean | Std. Dev |
|---|---|---|
| Observations | 698 | |
| | | |
| Income less than \$100K | 16.64% | |
| Income between \$100K to \$150K | 35.43% | |
| Income between \$150K to \$200K | 27.76% | |
| Income between \$200K to \$250K | 9.95% | |
| Income between \$250K to \$300K | 6.36% | |
| Income more than \$300K | 3.86% | |
| Income Type: Hourly | 4.48% | |
| Income Type: Salary | 71.73% | |
| Income Type: Profits from Practice | 23.79% | |
| | | |
| Hours Last Week | 50.19 | 13.75 |
| Weeks Worked | 47.48 | 4.63 |
| Full Time | 87.95% | |
| | | |
| Experience | 13.69 | 8.35 |
| Foreign Medical Graduate | 15.17% | |
| | | |
| Female | 30.83% | |
| | | |
| Practice Type: Solo/Two Physician | 31.82% | |
| Practice Type: Group | 46.27% | |
| Practive Type: Other | 21.91% | |
| | | |
| Large Metropolitan Area | 46.89% | |
| Small Metropolitan Area | 32.44% | |
| Non-Metropolitan Area | 20.67% | |

Notes: Sample of Family Practice Physicians in the Health Tracking Physician Survey of 2006 with non-missing income, starting medical practice in or before 2006. Income from medically related activities in 2006. Hours reported for medically-related activities. Income excludes in returns from investments in financial and medical capital. Experience defined as number of years since beginning medical practice. Full-time defined as more than 35 hours spent on medical activities and more than 40 weeks worked in 2006. Large Metropolitan Area has more than 1 million residents.

Table F.9: Income of Family Medicine Doctors

| Dependent Variable | Log Income from Practice | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| *Panel A: Interval Regression Estimates* | | | | | | |
| Experience | 0.0144* | 0.0124 | 0.0819** | 0.0117 | 0.0147* | 0.0126 |
| | (0.0070) | (0.0070) | (0.0256) | (0.0073) | (0.0069) | (0.0069) |
| Experience-squared | -0.0003 | -0.0004 | -0.0063** | -0.0004 | -0.0005 | -0.0004 |
| | (0.0002) | (0.0002) | (0.0024) | (0.0002) | (0.0002) | (0.0002) |
| Female | | -0.2617*** | -0.2621*** | -0.2121*** | -0.2581*** | -0.2759*** |
| | | (0.0352) | (0.0421) | (0.0372) | (0.0346) | (0.0347) |
| Foreign Medical Graduate | | | | | | -0.0446 |
| | | | | | | (0.0441) |
| Practice Type: Solo/Two Physician | | | | | | -0.1392*** |
| | | | | | | (0.0365) |
| Practice Type: Other | | | | | | 0.0062 |
| | | | | | | (0.0345) |
| Small Metropolitan Area | | | | | | 0.0544 |
| | | | | | | (0.0347) |
| Non-Metropolitan Area | | | | | | 0.0647 |
| | | | | | | (0.0398) |
| Contant | 11.7822*** | 11.9014*** | 11.7658*** | 11.9388*** | 11.8955*** | 11.9228*** |
| | (0.0413) | (0.0449) | (0.0596) | (0.0465) | (0.0433) | (0.0500) |
| Heteroskedascitiy by experience | | | | | Y | |
| Sample | | | Young | Full-time | | |
| Oberservations | 698 | 698 | 295 | 616 | 698 | 698 |
| Total Sample Weight | 60620 | 60620 | 25612 | 53318 | 60620 | 60620 |
| *Panel B: Estimated Distribution Statistics at Zero Experience* | | | | | | |
| Mean | 153660.74 | 148524.83 | 127612.97 | 157920.84 | 144746.55 | 146895.78 |
| Std. Dev. | 68769.35 | 63368.95 | 47622.36 | 65432.74 | 50911.87 | 61416.25 |

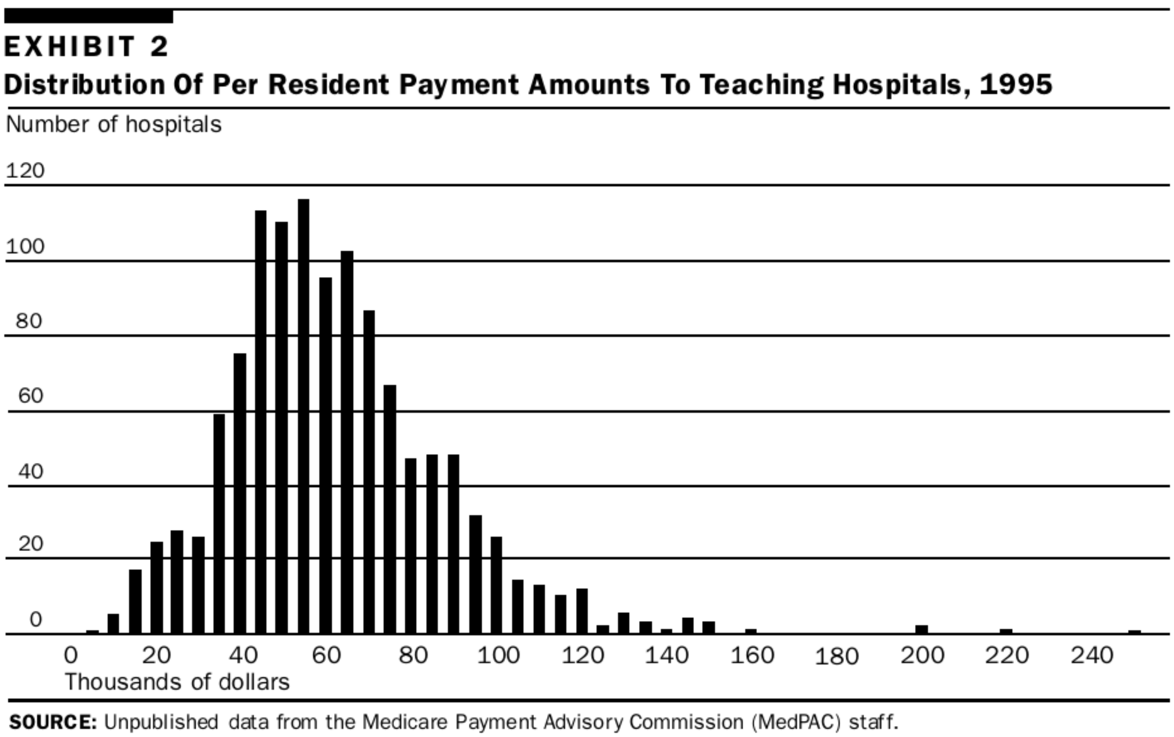
Notes: Interval regressions with normally distributed error. Baseline sample and characteristics as defined in Table F.8. Column (3) restricts to physicians with less than 10 years of experience. Column (4) estimates sigma as a quadratic function of experience. Earnings statistics in 2010 dollars, calculated at zero experience, mean gender and foreign graduate fractions observed in the resident population, and means of practice location and type characteristics (only for column (6)). Significance at 90% (*), 95% (**) and 99% (***) confidence.

Table G.10: Medicare Reimbursement Rates on Characteristics

| Dependent Variable | Log Competitor Reimbursements | | | Log Reimbursements | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Log Rent | -0.0057 | -0.0282 | -0.1746** | -0.0004 | -0.2023 | -0.1330 |
| | (0.0632) | (0.0737) | (0.0879) | (0.1219) | (0.1579) | (0.1973) |
| Log Wage Index | 0.3924*** | 0.3425*** | 0.0937 | 0.7497*** | 0.6509*** | 0.4406 |
| | (0.1036) | (0.1038) | (0.1042) | (0.1977) | (0.2042) | (0.2679) |
| Log Reimbursement | 0.1701*** | 0.1538*** | 0.1410*** | | | |
| | (0.0227) | (0.0221) | (0.0232) | | | |
| Location Characteristics | | Y | Y | | Y | Y |
| Small Cities (<3 mi in Population) | | | Y | | | Y |
| Observations | 3,441 | 3,441 | 2,407 | 3,441 | 3,441 | 2,407 |
| R-squared | 0.2335 | 0.2934 | 0.2719 | 0.3528 | 0.3731 | 0.3550 |

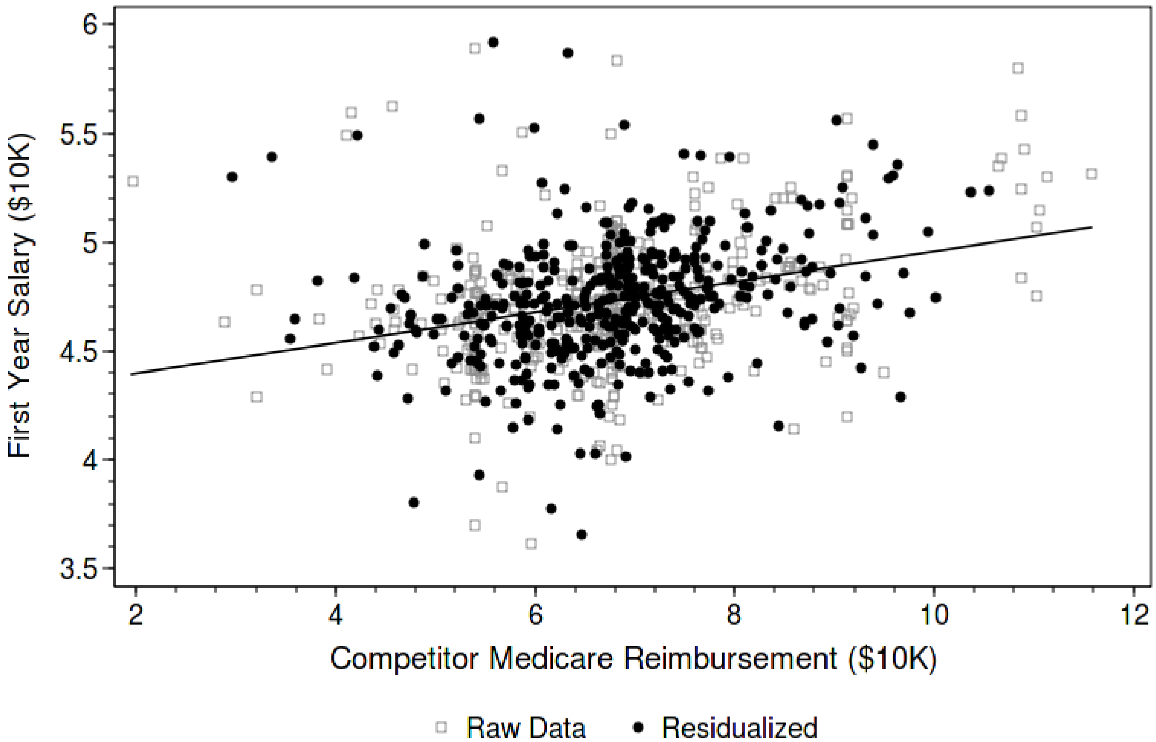
Notes: Linear regressions. Location characteristics include median age (county), log median household income (county), log total population (MSA/county), violent crime and property crime rates from FBI's Crime Statistics/UCR (25 mi radius weighted by 1/distance), dummies for no data in that radius and log college share (MSA/rest of state). All columns include a constant term, log # beds, log NIH Fund (Major), log NIH Fund (Minor), Log Case Mix Index, Program Type Dummies, Rural Program Dummy and dummies for programs with no NIH funding at major affiliates, for no NIH funding at minor affiliates, and a dummy for missing Medicare ID at program institutions. The Competitor Reimbursement is a weighted average of the Medicare primary care per resident amounts of institutions in the geographic area of a program other than the primary institutional affiliate of the program. Geographic area defined as in Medicare DGME payments: MSA/NECMA or Rest of State unless less than 3 other observations constitute the area, in which case the census division is used. See data appendix for description of variables and details on the construction of the reimbursement variables. For columns (4-6), a program's reimbursement rate is truncated below at $5,000 and a dummy for these 46 truncated observations is estimated as well. Standard errors clustered at the program level in parenthesis. Significance at 90% (*), 95% (**) and 99% (***) confidence.

Figure G.1: Distribution of Per Resident Amounts



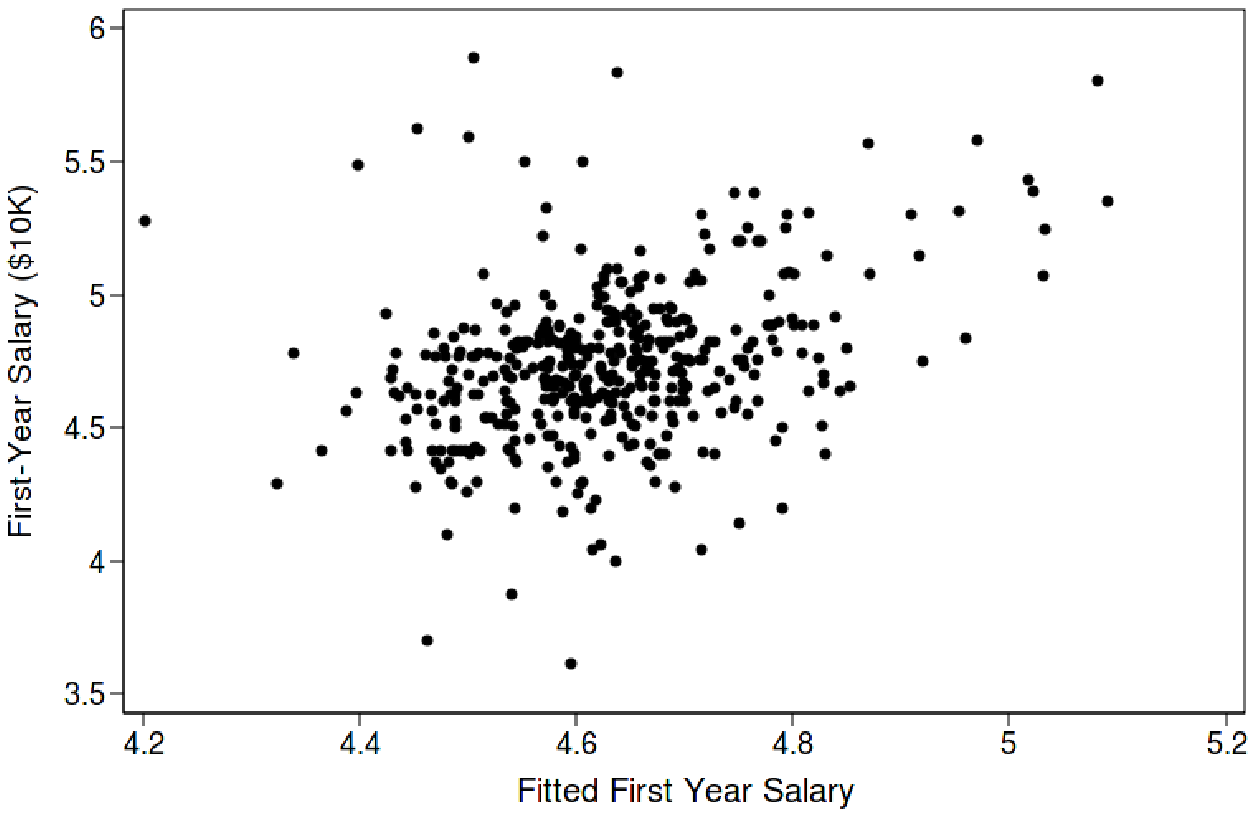

Notes: Secondary source from Newhouse and Wilensky (2001). A similar distribution can be roughly reproduced using the Medicare cost report data used in this study.

Figure G.2: Relationship Between Wages and Competitor Reimbursements



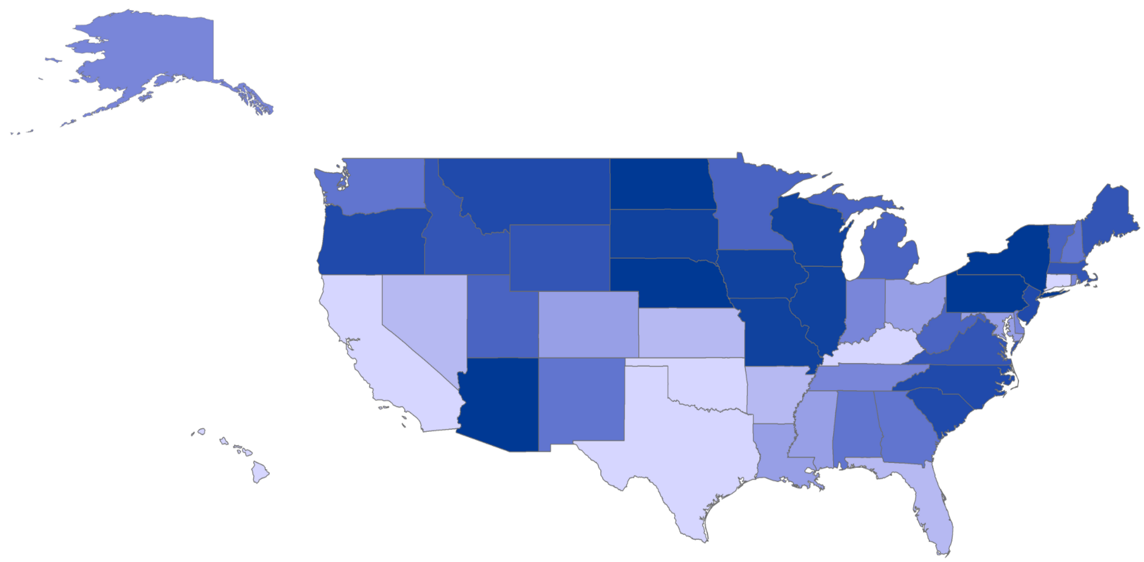

Notes: Sample restricted academic year 2010-2011. To construct the residualized scatter plot, I first regressed the X-axis and Y-axis variables on County Median Rent (Gross), Rural Program, Medicare Wage Index, Log NIH Fund (Major), Log NIH Fund (Minor), Log # Beds, Medicare Case-Mix Index and dummies 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 G.3: 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 G.4: 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.