Supplementary Material for
"Alternative Models for Moment Inequalities"
Simulated HMO-Hospital Contracting Model

Ariel Pakes and Robin Lee

1 Model Preliminaries

We analyze a market with M HMO plans, H hospitals, and N individuals. Each HMO k and hospital j possesses a vector of characteristics θ_k^M and θ_j^H respectively. Individuals are divided among R different demographic groups, where group r makes up share σ_r^R of the population and values HMO and hospital characteristics according to the coefficients β_r^M and β_r^H respectively. We assume any hospital can contract with any number of different HMOs, and similarly any HMO can contract with any number of hospitals. Denote by M_j the set of HMOs of which that hospital j is a member; similarly, let H_k represent the set of hospitals that are in HMO k's network of providers.

Note that either the set $\{M_j\}_{j\in H}$ or $\{H_k\}_{k\in M}$ can uniquely define a network or market structure. We choose to represent a network structure by the binary HxM matrix χ, where $\chi_{(j,k)} = 1$ iff hospital j is in HMO k's plan, and $\chi_{(j,k)} = 0$ otherwise.

1.1 Individual Choice

We assume every individual will be hospitalized with probability γ. If sick, in order to use a particular hospital $j \in H$, an individual needs to have enrolled in an HMO plan k with $j \in H_k$. Each HMO k charges a one-time premium p_k. There is also an outside option which provides the individual with necessary health care in the case of illness – the utility of this option is normalized to 0.

Let individual i be part of demographic group r. We define an individual i's utility from using hospital j as

$$u_{ij}^H = \theta_j^H \beta_r^H + \omega_{ij} \tag{1}$$

where ω is distributed iid Type I extreme value. From this formulation, we can define an individual i's utility from enrolling in a given HMO k that has a set of hospitals H_k:[1]

$$u_{ik}^M = \theta_k^M \beta_r^M - \alpha p_k + \gamma \left(ln(\sum_{j\in H_k} exp(\theta_j^H \beta_r^H)) \right) + \varepsilon_{im}$$

where ε is also distributed iid Type I extreme value.

With this linear utility function and distribution on error terms, we can calculate the the (expected)[2] share of the population that chooses HMO k given any particular network structure χ

[1]Note $E_\omega(\max_{h\in H_m}(u_{ij}^H)) = ln(\sum_{j\in H_m} \exp(\theta_j^H \beta_r^H))$

[2]For the purposes of discussion, we will assume all references to the shares determined by consumer choice are in fact expectations even when not explicitly staged – with a small N, the formulas given by the logit formulas will not

as follows:

$$\begin{aligned}\sigma_k^M &= \sum_{r\in R}\sigma_r^R \frac{\exp(\theta_k^M\beta_r^M - \alpha p_k + \gamma\ln(\sum_{j\in H_k}\exp(\theta_j^H\beta_r^H)))}{1+\sum_{l\in M}(\exp(\theta_l^M\beta_r^M - p_l + \gamma\ln(\sum_{g\in H_k}\exp(\theta_g^H\beta_r^H))))} \\ &= \sum_{r\in R}\sigma_r^R\tilde{\sigma}_{k,r}^M\end{aligned} \tag{2}$$

We use $\tilde{\sigma}_{k,r}^M$ to represent the share of demographic group r that chooses HMO plan k.

We can go even further, and define the demographic distribution of individuals within each HMO plan; that is, the share of people who use HMO plan k who are part of demographic group r:

$$\tilde{\sigma}_{r,k}^R = \frac{\sigma_r^R\tilde{\sigma}_{k,r}^M}{\sum_{s\in R}\sigma_s^R\tilde{\sigma}_{k,s}^M}$$

This allows us now to denote the share of HMO plan k's customers who actually will be sick and need to use hospital $j \in H_k$:

$$\sigma_{j,k}^H = \gamma\sum_{r\in R}\tilde{\sigma}_{r,k}^R\frac{\exp(\theta_j^H\beta_r^H)}{\sum_{l\in H_k}\exp(\theta_l^H\beta_r^H)} \tag{3}$$

Note that σ_k^M is a function of the entire network structure (χ) and premiums charged by all HMOs (P); also note that σ_k^H is just a function of HMO k's own hospital network H_k. For the rest of the exposition, when we use σ_k^M and σ_k^H, we will assume that a particular network structure and set of premiums is given.

1.2 Hospital and HMO Behavior

In our model we assume each hospital j offers each HMO k a take-it-or-leave-it contract which specifies a per-patient transfer of $T_{j,k}$ for each patient of HMO k that hospital j serves. We represent the set of per-patient transfers offered by all HMOs to each hospital as an HxM matrix T.

An HMO can reject or accept any contract offered to it. Thus, by accepting or rejecting these contracts, the HMOs directly determine the healthcare network structure of the market – i.e., which hospitals belong to which HMO plans. However, since hospitals choose the initial contract offers, they profoundly influence these eventual choices.

For hospital j, profits are given by the equation

$$\pi_j^H = \left[(\sum_{k\in M_h}N\sigma_k^M\sigma_{j,k}^H T_{j,k}) - c_j^H(\sum_{k\in M_h}N\sigma_k^M\sigma_{j,k}^H)\right] \tag{4}$$

c_j^H represents the average cost of serving each patient at hospital j. We assume that each hospital j has a "capacity constraint" Γ_j, and will face increasing costs if it serves more than Γ_j

be exact. However, for the purposes of this model, this distinction is irrelevant: given the timing of the model, when players make their decisions all choices are based on expectations; furthermore, when actually calculating profits, we will be allowing for "fractions" of people to perform actions. In other words, we are approximating a large population using a small value of N for notational purposes. Thus, a unit of N can be thought of as a "block" of people, and all previous definitions can be modified as implied by this interpretation.

patients. We assume that if hospital j serves x patients, its average costs per patient will be given by

$$c_j^H = \bar{c}_j^H + (x - \Gamma_j)^\phi 1_{x>\Gamma_j}$$

where $\bar{c}_j^H$ is an hospital specific constant, and ϕ represents the exponential parameter through which overflow patients affect average costs if the hospital is over capacity (denoted by the indicator function $1_{x>\Gamma_j}$).

For any HMO k, its profits are calculated as follows:

$$\begin{aligned} \pi_k^M &= N\sigma_k^M(p_k - c_k^M) - \sum_{j\in H_m} T_{j,k}N\sigma_k^M\sigma_{j,k}^H + \sum_{j\in H_k}\varepsilon_{j,k} \\ &= \tilde{\pi}_k^M + \sum_{j\in H_k}\varepsilon_{j,k} \end{aligned} \tag{5}$$

c_k^M represents the cost that each HMO incurs per patient under its plan; for the purposes of this model, we normalize this cost to 0 for all HMOs. Note that the profits of an HMO depends is influenced by the network structure of the industry, which in turn is determined not only by the set of transfers and premiums HMO k has, but those of its competitors as well.

In the case of asymmetric information, $\varepsilon_{j,k}$ represents the set of "Hospital-synergies" that a particular HMO k possesses for hospital j. I.e., ε are values for each hospital-HMO combination that enter additively into a particular HMO's profit function for each HMO in which that hospital is a member. We assume that each HMO knows only its own values of $\varepsilon_{j,\cdot}$; hospitals do not know any of the values of ε. However, all parties have a common prior over the distribution of all ε. For the purposes of our model, we assume that they are distributed as i.i.d. standard normals.

For this current version of the program, however, there is no asymmetric information. Thus, $\varepsilon = 0 \;\forall\; j,k$.

2 Timing of Actions

The model proceeds as a multistage game with observable actions between stages. In each stage, players choose their actions simultaneously. The solution concept is SPNE for the perfect information case, Perfect Bayesian Nash Equilibrium for the asymmetric info case.

2.1 Stage 1

Each hospital $j \in H$ chooses a set of functions $\{t_{j,k}\}_{k\in M}$ where $t_{j,k} : Z_+ \to \mathbb{R}$ which specifies a per-patient payment to hospital j for a given number of HMO k's patients served by that hospital. For now we restrict the set of contracts to be linear; specifically, that they be a fixed constant per-patient payment. I.e., each hospital $j \in H$ offers a set of contracts $T_j = (T_{j,1}, \ldots, T_{j,M}) \in \mathbb{R}_+^k$ to the hospitals.

2.2 Stage 2

At this stage each HMO k knows the full set of contracts $T = (T_1, \ldots, T_H)$ offered by all hospitals to all HMOs; in the asymmetric information case, each HMO knows only its own set of synergies $\{\varepsilon_{j,k}\}_{j\in H}$. Each HMO k chooses an action $\chi_k = (\chi_{1,k}, \ldots, \chi_{H,k}) \in \{0,1\}^H \equiv S$, where $\chi_{j,k} = 1$ iff HMO k accepts the contract offered by hospital j. Each HMO thus has a strategy space of size $\{2^H\}$. Note that the HMOs' actions at this stage define a network structure $\chi = (\chi_1, \ldots, \chi_H)$.

2.3 Stage 3

Each HMO k chooses a premium $p_k \in \mathbb{R}^+$ that it will charge each consumer that chooses to join its plan.

2.4 Stage 4

Each individual i of N total consumers chooses to enroll in an HMO plan, with the utility from choosing HMO k being $u^M_{i,k}$ defined in equation 1.1, or chooses to utilize the outside option, thereby deriving a utility of 0. The HMO choices by consumers thus determines the number of consumers enrolled in each HMO plan and the number of consumers that are treated at each hospital. HMO payoffs (π^M) and hospital payoffs (π^H) are realized.

3 Equilibrium Conditions

3.1 Perfect Information

A SPNE of this game will consist of a set of transfers offered by hospitals $T^* = (T^*_1, \ldots, T^*_H)$; HMO strategies $\{\chi_1, \ldots, \chi_M\}$ defined for all possible values of T; and a set of induced premiums $p(T,\chi) = (p_1(T,\chi), \ldots, p_M(T,\chi))$ for all possible realizations of (T,χ). These all must satisfy the following conditions:

- $p_k(T,\chi) \in \arg\max_{p_k} \pi^M_i(T,\chi,p_k,p_{-k}) \;\forall\, k \in M,\; \forall\, T,\chi$
- $\chi_k(T) \in \arg\max_{\chi_k} \pi^M_i(T,(\chi_k,\chi_{-k})] \;\forall\, k \in M,\; \forall\, T$
- $T^*_j \in \arg\max_{T_j} \pi^H_j((T_j,T^*_{-j}),\chi(T_j,T^*_{-j}),p) \;\forall\, j \in H$

3.2 Asymmetric information

A PBNE of this game will consist of a set of transfers offered by hospitals $T^* = (T^*_1, \ldots, T^*_H)$; a set of beliefs $\mu^*(T) = \{\mu^*_1(T), \ldots, \mu^*_M(T)\}$ over the space of all possible HMO strategies $\{\chi_1, \ldots, \chi_M\}$ defined for all possible values of T; and a set of induced premiums $p(T,\chi) = (p_1(T,\chi), \ldots, p_M(T,\chi))$ for all possible realizations of (T,χ). These all must satisfy the following conditions:

- $p_k(T,\chi) \in \arg\max_{p_k} \pi^M_i(T,\chi,(p_k,p_{-k})) \;\forall\, k \in M,\; \forall\, T,\chi$
- Given ε_k, $\chi_k(T) \in \arg\max_{\chi_k} E_{\mu^*_{-j}}[\pi^M_i(T,(\chi_k,\chi_{-k}))] \;\forall\, k \in M,\; \forall\, T$
- $\mu^*_k(\chi_k) = Pr(\varepsilon_{\cdot,k} | \chi_k \in \arg\max_{\chi_k} E_{\mu^*_{-k}}[\pi^M_k(T,(\chi_k,\chi_{-k})))]) \;\forall\, \chi_k \in S,\; \forall\, k \in M$
- $T^*_j \in \arg\max_{T_j} E_{\mu^*}[\pi^H_j((T_j,T^*_{-j}),\chi(T_j,T^*_{-j}),p)] \;\forall\, j \in H$

4 Solving the Model

We describe how an equilibrium is computed for a given set of parameter values by describing how each stage is computed. We proceed by backward induction – i.e., we describe how demand is realized given transfers, a network structure and set of premiums; we determine premiums given a network structure and set of transfers by being able to infer what realized demand will be; we determine the network structure given transfers by knowing what demand and premiums will be; and finally, we can figure out transfers since we can calculate for any set of transfers, what the implied network, premiums, and demand will be.

4.1 Stage 4: Individual Demand

At this stage, a set of premiums P, transfers T, and a network structure χ will be specified. Thus, consumer demand will be identified using equations 2 and 3, and expected profits can easily be calculated for all the Hospitals and HMOs with equations 4 and 5.

4.2 Stage 3: HMO Premiums

At this stage, a set of transfers T and network structure χ is specified.

In order to calculate a NE in premiums, we first calculate the best-response function for an HMO k given every other HMO has set premiums P_{-k}. Note we can re-express HMO profits as:

$$\begin{aligned}\pi_k^M &= N\sigma_k^M(p_k - c_k^M - \sum_{j\in H_m} T_{j,k}\sigma_{j,k}^H) \\ &= N\left(\sum_{r\in R}\sigma_r^R\frac{\Lambda_{k,r}}{\Delta_r}\right)(p_k - c_k^M - \sum_{j\in H_m} T_{j,k}\sigma_{j,k}^H)\end{aligned}$$

where

$$\begin{aligned}\Lambda_{k,r} &= \exp(\theta_k^M\beta_r^M - \alpha p_k + \gamma\ln(\sum_{j\in H_k}\exp(\theta_j^H\beta_r^H))) \\ \Delta_r &= 1 + \sum_{l\in M}(\exp(\theta_l^M\beta_r^M - \alpha p_l + \gamma\ln(\sum_{g\in H_k}\exp(\theta_g^H\beta_r^H))))\end{aligned}$$

Note only $\Lambda_{k,r}$, Δ_r and p_k are dependent on the choice of p_k.

Given the other HMO premiums, the optimal premium p_k^* for HMO k maximizes π_k^M. Assuming necessary second order conditions hold, this is equivalent to solving the following

$$\begin{aligned}0 &= \frac{\partial\pi_k^M(T,\chi,p_k^*,P_{-k})}{\partial p_k} \\ &= \frac{\partial\sigma_k^M}{\partial p_k}(p_k^* - c_k^M - \sum_{j\in H_m} T_{j,k}\sigma_{j,k}^H) + \sigma_k^M \qquad (6)\end{aligned}$$

where

$$\frac{\partial\sigma_k^M}{\partial p_k} = \sum_{r\in R}\sigma_r^R\frac{\partial\tilde{\sigma}_{k,r}^M}{\partial p_k} = \sum_{r\in R}\sigma_r^R\frac{\partial\frac{\Lambda_{k,r}}{\Delta_r}}{\partial p_k} = \sum_{r\in R}\sigma_r^R\frac{\Lambda_{k,r}(\Lambda_{k,r} - \Delta_r)}{\Delta_r^2}$$

since

$$\frac{\partial\Delta_r}{\partial p_k} = -\Lambda_{k,r}$$

Consequently, we define the best response premium function for HMO k as $BR_k(p_{-k}) = p_k^*$ as defined in the equation 6.

Implementation

Since the best response function is not analytically invertible, we can still numerically approximate what it would be.[3] Thus, starting with an initial vector of premiums charged by the HMOs, we

[3]Given p_{-k}, we plug the vector (p_k, p_{-k}) into equation 6 starting with $p_k = 0$. Since HMO profits are increasing in its own premium from $p_k = 0$, this FOC will be positive for $p_k = 0$. We then increase p_k by a fixed amount (.01 in the program) and keep doing so until equation 6 is negative. This is our approximation for $p_k = BR(p_{-k})$. Though it is true that this value of p_k may technically only be a local maximum for profits, it is unlikely that a different global maximum exists.

update each HMO's premium to be a best response to the other HMOs. We iterate until a fixed point is found. This vector p^* is a NE in the premium setting stage.

4.3 Stage 2: HMO Choices

At this stage, hospitals will have offered the set of per-patient transfers $T \equiv \{T_1, \ldots, T_H\}$.

4.3.1 Perfect Information Case

We start by assuming that all HMOs contract with all hospitals. We then allow the first HMO to choose among all of its possible options (e.g., contracting with any subset of hospitals) holding the other HMO actions fixed, and select its best response.[4] We then move to the second HMO, and repeat. We continue iterating, allowing each HMO in turn to make its best response to the current iteration's network structure, until we 1) converge to a fixed point or 2) cycle. If a fixed point is found, it is a NE of this stage. If the process cycles, then we assume that no one contracts with anyone (although this has not occurred in the 2x2 case).

4.3.2 Asymmetric Information Case

Assume that a particular HMO k has beliefs μ_{-k} over χ_{-k}, i.e. the network structures its competitor hospitals will choose. Then, since this particular HMO k knows its own set of ε_k's, it will choose the appropriate $\chi_k \in S$ such that

$$\begin{aligned}
\chi_k &\in \arg\max_{\chi_k} E_{\mu_{-k}}[\pi_k^M(T, \chi_k, \chi_{-k})] \qquad (7)\\
&\in \arg\max_{\chi_k} \left[\sum_{\chi_{-k} \in S_{-k}} \mu_{-k}(\chi_{-k}) \left(\sum_j N\sigma_k^M(\chi_k, \chi_{-k})[(p_k - c_k^M) - \sum_{j \in H_k} \sigma_{j,k}^H(\chi_k, \chi_{-k}) T_{j,k}] \right) \right] + \chi_j \varepsilon_j
\end{aligned}$$

To HMO k's competitors, since the distribution of ε_k is known, they can consequently construct the distribution μ_k over S for any given μ_{-k} by calculating the probability of HMO k realizing a vector ε_k which induces a particular choice $\chi_k \in S$. An equilibrium set of beliefs μ^* is such that $\mu_k^* = \mu_k(\mu_{-k}^*)$ $\forall$ k – i.e., the equilibrium beliefs μ_k^* over which strategy hospital k will choose coincide with the actual ex ante probabilities that hospital k will choose these strategies.

These equilibrium beliefs can be found via an iterative fixed point method – i.e., start with an arbitrary set of beliefs μ^0. For each HMO k, form $\mu_k^1 = \mu_k(\mu_{-k}^0)$. Let $\mu^1 = \{\mu_1^1, \ldots, \mu_M^1\}$, and keep iterating until these beliefs converge: $||\mu^n, \mu^{n+1}|| < \kappa$ for some specified norm and tolerance κ.

4.4 Stage 1: Hospital Transfers

For any given set of transfers T, the procedure in Stage 2 should be able to generate either 1) the network structure χ induced by T, or 2) equilibrium beliefs μ^* over the space of possible χ's. Thus it becomes relatively straight forward to compute an hospital's expected profits given a set

[4] Note that when it calculates the value of each of its options, a new network structure is implied and hence a new set of premiums and induced demand will be computed. Hence, the nested nature of this approach.

of transfers offered by all hospitals – in the asymmetric case it is as follows:

$$\begin{aligned} \pi_j^H(T) &= E_{\mu^*}[\pi_j^H(T,\chi)] \\ &= \sum_{\chi \in S^M} \mu^*(\chi) \left(\sum_k [N\sigma_k^M \sigma_{j,k}^H T_{j,k}] - c_j^H (\sum_k N\sigma_k^M \sigma_{j,k}^H T_{j,k} \right) \end{aligned}$$

We proceed via an iterative best-response dynamic.

Implementation

We proceed by discretizing the space of allowable transfers per patient each hospital can make into units τ.

Note that HMO k can spend p_k/γ per patient before it starts to lose money on a particular hospital contract. Thus, the highest per-patient transfers each hospital can make to a particular HMO k be the value $\bar{t}_k = \bar{Z}_k \tau$, where $\bar{Z}_k \in \mathbb{Z}^+$ is the largest integer such that $\bar{Z}_k \tau \leq p_k/\gamma$. We find the largest value of $\bar{Z}_k$ of all k – which we call $\bar{Z}$. Thus for each hospital j, we define the set of strategies that it can use for each HMO k as $\mathbb{A} \equiv \{0, 1, 2, \ldots, \bar{Z}\}$ where the choice of $z_{j,k} \in \mathbb{A}$ determines $T_{j,k} \equiv z_{j,k}\tau$. [5]

Given an arbitrary starting set of transfers T^0, a hospital j's best response $T_j^*(T_{-j}^0)$ can be found by searching over all possible strategies $\times_j \mathbb{A}$, and determining which one yields the highest profits for the given set T_{-j}^0.[6] Given this procedure for calculating any hospital j's best response to transfers T_{-j}, we can keep cycling through hospitals and replacing the set of transfers with best responses until we converge to a set of transfers where no hospital will wish to unilaterally deviate. If such a set of transfers is found, it is a pure strategy NE.

This NE T^* along with the induced network structure $\chi(T^*)$ and premiums $p(T^*, \chi(T^*)$ is a SPNE of this game.

5 Parameters

We first focus on an example with 2 HMOs and 2 Hospitals. Units, unless otherwise specified, are in thousands.

5.1 Demographic Characteristics

People are hospitalized with probability $\gamma = .075$. We assume that markets should be, in expectation over the aggregate, not over-capacity. Market size is distributed normally with a mean of 600, standard deviation 300, and a minimum value of 100; thus, if the mean number of people enroll in an HMO plan, 40 patients will need to be served.

There are three demographic groups, each of whose share of the overall population is found by drawing a uniform random for each group, and then appropriately calculating the relative weights. Each group's preference for any particular hospital or HMO β_r is drawn uniformly from the range $[.9, 1.1]$.

Disutility of premium prices $\alpha = 1.5$.

[5] So far, this upper bound has not been reached with given parameter values; however, if it is ever reached as a hospital's out-of-equilibrium best response in the middle of a computation, the upper bound is automatically raised. This is mainly a computational issue – since each time a hospital needs to determine its best response the entire space of transfers much be searched, the smaller this space the faster the computation.

[6] We begin by assuming that starting transfers offered by each hospital $T_{j,k}^0 = z_j \tau \ \forall \ k$, where z_j is the smallest element in $\mathbb{A}$ such that $z_j \tau > c_j^H$.

5.2 HMO Characteristics

HMO per-patient costs c_k^M are normally distributed with mean .75 and standard deviation .25. HMO quality θ_k^M is distributed normally with mean 2.5 and standard deviation .25. It is correlated with costs by a value of $\rho^M = .5$.

5.3 Hospital Characteristics

Hospital quality θ_j^H for each hospital are normals with mean $\mu_{\theta^H} = 25$ and standard deviation $\sigma_{\theta^H} = .5$. Hospital constant marginal costs $\bar{c}_j^H$ are gaussian normals with mean $\mu_{\bar{c}^H} = 12$ and standard deviation $\mu_{\bar{c}^H} = 9$, with a minimum of 1. Costs and hospital quality index for a particular hospital j are correlated by a value of $\rho^H = .5$. For each pair, these variables are generated first by creating two correlated standard normal random variables, and then appropriately transforming them with the correct mean and standard deviation.

We assume that hospital capacity Γ_j is i.i.d. normal, with mean $\mu_\Gamma = 25$ and standard deviation $\sigma_\Gamma = 10$, with a minimum value of 1. Marginal costs increase exponentially in the number over capacity by the value $\phi = 1.5$.