Network Formation and Bargaining in Vertical Markets: The Case of Narrow Networks in Health Insurance

Soheil Ghili∗

February 8, 2018

Abstract

“Network Adequacy Regulations” intend to help consumers by forcing “narrow-network” insurance plans to include more hospitals. But they can also give hospitals excessive bargaining leverage, leading to increased reimbursement-rates and premiums. To study this, I develop and estimate a model of network formation and bargaining between hospitals and insurers. Crucially, my bargaining formulation allows insurers to threaten to replace an in-network hospital with an out-of-network one. Applied to a health-insurance market in Massachusetts, my model predicts that regulations mandating large minimum network-sizes can raise the prices substantially. Also, surprisingly, network-adequacy regulations can cause “broad-network” plans to downsize.

Insurers increasingly use “narrow network” plan designs in the health insurance exchanges established by the Affordable Care Act (ACA). These plans offer their enrollees a small set of “in-network” hospitals to choose from, and if the enrollee decides to go “out of network,” s/he will have to

∗Yale University, New Haven CT, email: soheil.ghili@yale.edu. I am grateful to Igal Hendel, David Dranove, Peter Klibanoff, and William Rogerson. I also thank seminar participants at Northwestern, Berkeley, Rochester, MIT, Maryland, Wisconsin, Yale, FTC, Penn, and Stanford. All errors are mine. I do not have any financial or other material interest related to this research. I obtained IRB approval for this project from Northwestern University.
pay all or most of the healthcare expenses out of pocket.\footnote{According to McKinsey & Company, about 60% of the ACA exchange plans cover less than 70% of local hospitals where their plans are offered, and about 20% of the plans cover less than 30% of local hospitals. Networks are expected to further narrow in 2017.} Narrow hospital networks have received vast and often unfavorable media attention. Some provider- as well as consumer-advocacy groups have filed lawsuits against both insurers and the federal government regarding narrow networks. They have also pressed the administrators of the healthcare exchanges, mainly Centers for Medicare and Medicaid Services (CMS), for “network adequacy regulations,” which are regulations that would force minimum mandated network sizes on insurers in order to expand narrower networks and increase consumers’ access to hospitals.\footnote{Narrow networks are not confined to hospitals. The same is true of other types of providers, such as physician groups and prescription drugs. My focus in this paper is on hospitals.}

However, the impacts of network adequacy regulations are not confined to hospital access. They also can have price consequences. Under such regulations, insurers cannot drop a hospital out of network if that action would take them below the minimum mandated network size. Taking advantage of this restriction faced by insurers, hospitals may gain further bargaining leverage and successfully negotiate higher reimbursement rates with insurers for the care they provide to patients. Insurers can, in turn, respond to the increased rates by raising their monthly premiums and passing the extra cost on to consumers.

The objective of this paper is to quantify the price consequences of network adequacy regulations. I do this by developing an empirically estimable model of insurer-hospital markets with two crucial features. First, I endogenize how insurers (i) form hospital networks, (ii) bargain with hospitals over rates, and (iii) set premiums. Second, my formulation of the bargaining procedure allows insurers to try to negotiate lower rates with hospitals not only by threatening to drop them from the hospital network, but also by threatening to replace them with currently out-of-network hospitals. I leverage detailed medical claims and insurance plan enrollments data to structurally estimate the model for the CommCare market, a health insurance exchange in Massachusetts, predating the ACA but similar to the ACA exchanges in many respects. I then use the estimated model to simulate a range of net-
work adequacy regulations and study how they impact hospital networks, negotiated rates, and premiums. I identify and analyze multiple economic forces that govern how narrow network plans respond to the regulation as well as how broader network plans, which are not directly affected by the regulation, respond to the responses of narrower network ones.

My model consists of a game with two main steps. In the first step, equilibrium hospital networks and reimbursement fees are simultaneously determined. In the second step, insurers set premiums la Bertrand. To model the first step, I impose network stability conditions and bargaining conditions on the equilibrium hospital networks and reimbursements. For network stability, I adapt pairwise stability conditions from Jackson and Wolinsky (1996). I require that at the equilibrium, no insurer or hospital can strictly profit either from dropping a bilateral contract or from adding a new bilateral contract that the other party would be willing to sign.

My bargaining conditions offer a substantial advantage over standard approaches to the modeling of bargaining in vertical markets, by accounting for an economic force that arises in settings in which the bargaining does not take place on a fixed network, but rather affects and is affected by network formation. The standard model of bargaining on a network in the empirical industrial organization literature is called Nash-in-Nash (NiN henceforth), and is based on Horn and Wolinsky (1988). NiN assumes that the reimbursement rate negotiated between hospital $i$ and insurer $j$ is the outcome of a Nash Bargaining procedure (la Binmore, Rubinstein and Wolinsky (1986)) between the two, taking as given the rest of the network structure and reimbursements. The construction of the Nash Bargaining formulation roughly implies that hospital $i$ can charge a high rate if insurer $j$ would lose a substantial amount of profit from leaving hospital $i$ out of its hospital network, but not modifying the network otherwise. This is the case under NiN even if insurer $j$ would lose little or no profit, or would gain some profit, from leaving hospital $i$ out and instead accepting a competing offer made by a currently-out-of-network hospital $i'$ that is trying to outbid hospital $i$ and replace it in the network, in a negotiation process simultaneous with the negotiation between $i$ and $j$. In other words, the negotiated reimbursement rate predicted by NiN is not affected at all by which hospitals are outside the network of insurer $j$ and how close substitutes they are to hospital $i$.

To capture insurers’ ability to use such out-of-network hospitals as “threats of replacement” is critical in the context of studying network adequacy regulations. For an insurer that is just meeting the minimum network size
mandated by a network adequacy regulation, dropping hospitals without replacement is not an option. Thus, for such insurers, threats of replacement are the only bargaining chip that can help to keep reimbursement rates low. Not accounting for these threats, NiN would predict unrealistically high charges by hospitals to insurers bound by such regulations. My bargaining formulation, instead, assumes that the rate negotiated between hospital $i$ and insurer $j$ is equal to the Nash Bargaining rate only if under that rate, the insurer cannot strictly profit from replacing $i$ with a currently out-of-network $i'$ at the lowest rate that $i'$ would accept. If, however, under the Nash Bargaining rate, insurer $j$ can strictly profit from such replacement, my bargaining formulation assumes that hospital $i$ brings its rate down to a level that would make the insurer indifferent between keeping hospital $i$ and replacing it with hospital $i'$.

To provide non-cooperative support, I extend the Rubinstein alternating offers game from bi-lateral to $2 \times 1$ (i.e., two upstream firms and one downstream) and show that my formulation can be sustained as an outcome of a Subgame Perfect Nash Equilibrium.

My empirical analysis has several steps, in which I estimate demand and cost functions for insurers and hospitals. First, I use detailed medical claims data on CommCare to estimate a model of hospital choice, backing out perceived hospital qualities by patients for different diagnoses as well as patients' disutility from travel. Based on these estimates, I construct network-expected-utility measures for each CommCare plan across different enrollee demographic groups. I then use the network-expected-utility measures along with data on other plan characteristics and on CommCare enrollments to estimate a demand model for insurance plans. I estimate marginal costs of inpatient care to hospitals using hospital cost reports, and to insurers, using the observed payments in my CommCare medical claims data. Finally, I combine all of these estimates and impose the structure of my model to back out relative bargaining powers for insurers and hospitals, as well as insurers’ fixed and variable non-inpatient costs.

I use the estimated model to simulate a range of network adequacy regula-

---

4 That is, hospital $i$ will charge the highest rate that deters threats of replacement by the insurer. Of course, if such a rate is so low that hospital $i$ would rather drop the contract, then pairwise stability is violated and current network structure cannot be part of the equilibrium configuration. Hence, the replacement must take place. See section 2 for more details.

5 Sources of non-inpatient costs can be outpatient care, doctor visits outside of hospitals, prescription drugs, and administrative costs.
tions. I focus on the Greater Boston Area and the 2011 fiscal year. I consider regulations in the form of forcing all insurers in the market to cover at least $X\%$ of the hospital systems present in the market. A higher $X$, hence, means a tighter regulation. I simulate the regulation over a range of values for $X$, and examine the responses of the hospital networks, reimbursement rates, and premiums.

My main finding is that under tight regulations, with $X > 85\%$, affected insurers (i.e., those that are forced by the regulation to expand their networks) experience large increases in the average reimbursement rates they pay to their in-network hospitals. Some of these insurers respond by raising their monthly premiums. For instance, Celticare, a CommCare plan that covered four out of the 16 hospital systems in 2011, is predicted to pay about $4,000 (or 28\%) more to its in-network hospitals per average hospital admission, when it is forced to cover at least 14 hospital systems. Celticare responds by raising its premium from $404/month to $425/month. One of the key driving forces behind this result is a sorting effect: the additional hospitals that insurers include in response to the regulation are those with the lowest cost and/or in highest demand. In other words, they are the “best” hospitals. This would leave the insurer with a pool of out-of-network hospitals comprising the “worst” $(100 - X)\%$ of the hospitals in the market. Thus, with high $X$, the threats of replacement that the insurer can make using its weakened out-of-network pool only become credible when the in-network hospitals, in particular the best ones among them, charge high rates.

Besides the main result, my simulations also point to other interesting predictions about the functioning of health insurance markets under network adequacy regulations. I find that for lower ranges of $X$, in which the sorting effect that weakens the out-of-network pools of insurers is less stringent, network adequacy regulations may in fact lower the average reimbursement rates paid by some insurers. This happens when the hospitals added in response to the regulation have lower marginal costs of providing care than

---

6Others do not raise premiums due to a premium cap which was imposed by the Massachusetts Health Connector, which administered CommCare.

7The premium could rise even further if there were no regulated premium cap at $425/month in CommCare. I will discuss these premium regulations in section 3.2.

8On top of this sorting effect, there are also more complex mechanisms that lead to rapid accelerations in the average reimbursement rate per admission that Celticare pays to its in-network hospitals and in Celticare’s premium. Those are discussed in section 5.1.
the insurer’s pre-regulation hospitals. I also find that even though they expand the narrower hospital networks, in some cases network adequacy regulations shrink the broader ones. This happens because some hospitals that get to charge higher rates to a narrow network plan that is constrained by the regulation, are encouraged to abandon some of the broader hospital networks in order to steer their more loyal patients to the plan that pays those hospitals more for treating the patients. In section 5, I discuss these findings in more detail, conduct welfare analysis for CommCare, and explain some of the differences between CommCare and the ACA exchanges that may lead to different responses between the two markets to network adequacy regulations.9

The framework I develop in this paper applies beyond network adequacy regulations. It can be applied to problems in which the interaction between network formation and bargaining is key. For example, it may be used to study how hospital-networks would respond to increased hospital bargaining power due to hospital consolidation. It may also be applied to other two-sided markets such as the market between TV channels and Cable companies.

The rest of the paper is organized as follows. Section 1 reviews the related literature. Section 2 sets up the model. Section 3.2 describes the CommCare market. Section 4 explains the estimation procedure and estimation results. Section 5 presents and interprets the counterfactual simulation results for CommCare and discusses potential differences from the ACA exchanges in terms of consumer-welfare implications. Section 6 concludes.

1 Related Literature

This paper can be considered part of the expanding body of literature in economics on issues related to the implementation of the Affordable Care Act. Recent papers have examined issues like subsidy schemes (Tebaldi (2015)), tiered networks (Prager (2015)), market competition and premiums (Dafny, Gruber and Ody (2015)), and selection (Hackmann, Kolstad and Kowalski (2015); Shepard (2015)). Some papers have studied issues directly related to narrow networks by comparing the health outcomes between enrollees of

---

9In the case of CommCare, even stringent regulations improve consumer welfare due to the very market-specific fact that in CommCare, premiums faced a mandatory cap set by the government. Thus, the increased reimbursement rates resulting from network adequacy regulations did not fully pass through to customers.
narrow and broad network plans (Gruber and McKnight (2014)), examining the relationship between network breadth and premium (Dafny, Hendel and Wilson (2015)), and estimating how much different groups of consumers value network breadth (Ericson and Starc (2014)). My paper takes another step by studying the consequences of regulating the networks for insurer-hospital bargaining and insurers’ premiums.

From a methodological standpoint, this paper also makes two contributions to the empirical analysis of vertical markets (e.g., Crawford and Yurukoglu (2012); Crawford et al. (2015); Gowrisankaran, Nevo and Town (2013); Ho and Lee (2017); Prager (2015); Pakes (2010); Ho (2009, 2006); Lee and Fong (2013)). First, I develop and empirically estimate a model that fully endogenizes network formation, bargaining, and downstream price setting. Second, by adding threats of replacement to NiN bargaining, I allow the upstream firms that a downstream firm excludes to impact its bargaining with those that it includes.

Also close in spirit to this paper are Ho and Lee (2017) and Liebman (2017). They also extend NiN and allow for endogenous networks to study network adequacy regulations. This paper differs in that in addition, it provides a general estimation approach for rationalizing the observed network and reimbursement rates. Also, computationally, my model enables me to analyze more hospitals and, especially, multiple insurers. This helps to study the indirect response of broad network plans to the regulation, which I will show is a first-order thing to consider in policy design.

Another difference is in the network-formation models. I use pairwise stability which implies any contract with gains-from-trade is signed at the equilibrium. Ho and Lee (2017) and Liebman (2017) assume the insurer can commit to excluding a certain number (or set) of hospitals (in spite of gains from trade) to use them as threats of replacement. A different form of insurer-commitment to narrow networks, based on volume-steering, has been mentioned in the literature: the insurer forms a narrow network and steers patient volume to in-network hospitals; and in exchange for that, those hospitals charge lower reimbursement rates (Polsky, Cidav and Swanson (2016), AIS (2014), McKinsey (2013), etc). Volume-steering can be implemented in my model by simply assuming the contracts are two-part tariffs rather than linear.\(^{10}\)\(^{11}\) These two types of commitment have different implications.

\(^{10}\) Nevertheless, I model the contracts as linear. The reasoning is discussed in section 3.1.

\(^{11}\) In the insurance industry, these non-linear contracts are often implemented relation-
Unlike volume-steering, Ho and Lee (2017a) and Liebman (2017) imply that an insurer can get a possibly large discount from hospital $u_i$ by excluding $u_i'$ even if the two are not competitors (e.g., when they are geographically far from each other; or if specialty hospitals are studied, when $u_i$ is a cancer hospital and $u_i'$ is a children hospital.) This can lead to predicting more negative consumer-welfare implications for narrow networks.

2 Model

2.1 Basic Setup and Notations

My model captures interactions among three types of players in a vertical market: $m$ hospitals (upstream firms, denoted $u_i$), $n$ insurers (downstream firms, denoted $d_j$), and consumers (who are enrollees for insurers and patients for hospitals). The state of the market, or the “market outcome,” is described by three important elements: network structure, denoted $G$; reimbursement rates, denoted $T$; and premiums, denoted $P$. Figure 1 exhibits the schematic and matrix representations of $G$, $T$ and $P$ through an example with three hospitals and two insurers.

Network $G$ represents who contracts with whom, or equivalently, which hospitals are covered by each insurer. It can be represented graphically (as in figure 1) or using a matrix of zeros and ones. Reimbursements matrix $T$ is also represented in a matrix form with elements $t_{ij}$. As is clear from the figure, for

ally. That is, the in-network hospital offers a low rate with the expectation that the insurer will exclude its competitors. If the insurer fails to do so, the hospital will not offer the discount next period.

Figure 1: Schematic and matrix representations of the state of the market
every inactive link \( g_{ij} = 0 \) in network \( G \), the corresponding reimbursement rate is null: \( t_{ij} = \emptyset \). For an active link \( g_{ij} = 1 \), the corresponding \( t_{ij} \) is interpreted as follows: insurer \( d_j \) reimburses hospital \( u_i \) with \( t_{ij} \) dollars for every unit of healthcare provided to an enrollee of \( d_j \) at hospital \( u_i \). Also, each element \( p_j \) of \( P \) is the amount each enrollee of insurer \( d_j \) pays to that insurer per month.

I assume that the expected profit to each firm in the market is solely a function of the market outcome. That is, the profit functions are in the form of \( \pi_{u_i}(G, T, P) \) and \( \pi_{d_j}(G, T, P) \) for all \( u_i \) and \( d_j \). The exact nature of the profit functions arises from demand and cost functions in the market, which I will turn to in later sections. My model of the market takes the profit functions as primitives and predicts what market outcome(s) \((G^*, T^*, P^*)\) will arise at the equilibrium.

This model consists of a sequential game with four main steps, as depicted in figure 2. In step 1, hospitals and insurers will engage in a “network-formation and bargaining” game. The outcome of this step is the equilibrium network and reimbursements \((G^*, T^*)\). The structure of this step will be discussed in section 2.2. In step 2, insurers engage in a premium-setting game la Bertrand, taking into account the outcome of the previous step, \((G^*, T^*)\). The outcome of this step is the set of equilibrium premiums \(P^*\). In the third step, consumers decide which insurance plan to buy in the market. Finally, in the fourth step, if they get sick and need hospitalization, they decide which hospital within their plan’s network of providers (as specified by \(G^*\)) to visit for treatment.

Step 2 in this game is fairly straightforward. Steps 3 and 4 are explained in more details in the estimation section. In the remainder of this section, I’ll focus on step 1, where network formation and bargaining are endogenously captured.

### 2.2 Network-Formation and Bargaining Game

I combine two sets of conditions in order to characterize the equilibrium pair of network structure and reimbursements \((G^*, T^*)\). The first set consists of network-stability conditions, which roughly require that no unilateral or joint deviation (among a pre-specified set of deviations) from \(G^*\) be able to strictly pareto-improve the profits to the firms participating in the deviation. The second set of conditions consists of bargaining conditions which determine, for all \( ij \) with \( g_{ij}^* = 1 \), what \( t_{ij}^* \) the negotiation between hospital \( u_i \) and
The rest of this subsection is organized as follows. First I develop a base model with very simple network stability and bargaining conditions. I then demonstrate, using an example, that the base model is inadequate for capturing competition among hospitals for inclusion in insurers’ hospital networks, which is crucial to my empirical analysis in this paper. I then modify the bargaining conditions to capture this force. I close this subsection by discussing my model’s advantages over some other potential alternatives.

2.2.1 Base Model

I start by introducing the network stability conditions, which are adapted with small modifications from the notion of pairwise stability in Jackson and Wolinsky (1996).

**Definition 1** The network-reimbursements pair \((G^*, T^*)\) satisfies “pairwise network-stability” if the following hold:

(i) For \(\forall g_{ij} = 1\), neither \(u_i\) nor \(d_j\) can strictly profit by unilaterally severing the link \(g_{ij}\). That is:

\[
\pi_{u_i}(G^*, T^*) \geq \pi_{u_i}(G^*_{-ij}, T^*_{-ij})
\]

(1)

\[
\pi_{d_j}(G^*, T^*) \geq \pi_{d_j}(G^*_{-ij}, T^*_{-ij})
\]

(2)
(ii) For \( g_{ij} = 0 \), there is no contract that \( u_i \) and \( d_j \) can sign that will yield a strict pareto improvement in their profits. That is:

\[
\exists t_{ij} \in \mathbb{R} \text{ s.t. } \pi_{u_i} \left( G^*_{+,ij}, T^*_{+,t_{ij}} \right) \geq \pi_{u_i} \left( G^*, T^* \right) \& \pi_{d_j} \left( G^*_{+,ij}, T^*_{+,t_{ij}} \right) \geq \pi_{d_j} \left( G^*, T^* \right) \tag{3}
\]

with at least one inequality holding strictly.\(^{[12]}\)

where \( (G^*_{-,ij}, T^*_{-,t_{ij}}) \) is constructed from \( (G^*, T^*) \) by switching \( g_{ij} \) from 1 to 0 and \( t_{ij} \) to \( \emptyset \); and \( (G^*_{+,ij}, T^*_{+,t_{ij}}) \) is constructed by doing the exact inverse.

Also, in this section the notation on \( P^* \) has been suppressed, as it has been assumed that \( P^* \) is anticipated by firms at the first stage of the game.

The bargaining conditions for the base model come from the NiN idea in \cite{HornWolinsky1988}.

**Definition 2** The network-reimbursement pairs \( (G^*, T^*) \) satisfies “NiN bargaining” relative to the “bargaining parameters matrix” \( \gamma_{m \times n} \in [0,1]^{m \times n} \) if, for \( g_{ij} = 1 \), we have \( t_{ij}^* = t_{ij}^{NB} (G, T, \gamma_{ij}) \), where \( t_{ij}^{NB} (G, T, \gamma_{ij}) \) is defined as:

\[
\arg \max_{\tilde{t} \in \mathbb{R}} \left[ \pi_{u_i} \left( G^*, (\tilde{t}, T^*_{-,t_{ij}}) \right) - \pi_{u_i} \left( G^*_{-,t_{ij}}, T^*_{-,t_{ij}} \right) \right]^{\gamma_{ij}} \times \left[ \pi_{d_j} \left( G^*, (\tilde{t}, T^*_{-,t_{ij}}) \right) - \pi_{d_j} \left( G^*_{-,t_{ij}}, T^*_{-,t_{ij}} \right) \right]^{1-\gamma_{ij}} \tag{4}
\]

where \( (\tilde{t}, T^*_{-,t_{ij}}) \) is constructed from \( T^* \) by substituting \( \tilde{t} \) for its \( ij \) element.

The rough intuition behind the NiN conditions is that at the equilibrium, every \( ij \) pair is taking the rest of \( (G^*, T^*) \) as fixed, and negotiates over how to divide the total surplus created due to the presence of \( g_{ij} \). The value \( t_{ij}^* \) is the one that divides this total surplus between \( u_i \) and \( d_j \) based on their respective bargaining parameters \( \gamma_{ij} \) and \( 1 - \gamma_{ij} \).\(^{[13]}\)

\(^{[12]}\)Pairwise stability can in principle be considered part of the NiN framework. However, in this paper, I introduce it as a separate set of conditions since in the literature, the term NiN is mostly used only for surplus division under a fixed network. Either way, this paper does not claim a theoretical contribution in introducing pairwise stability. It adds value by using them in estimation and counterfactual simulations (all \( ij \) links can endogenously change in the equilibrium.)

\(^{[13]}\)What the Nash Bargaining formulation formally does (which is to maximize a weighted product of the surpluses made by \( u_i \) and \( d_j \), where the corresponding weights are \( \gamma_{ij} \) and \( 1 - \gamma_{ij} \)), does not always lead to dividing the total surplus between the two according to shares of \( \gamma_{ij} \) and \( 1 - \gamma_{ij} \). Nevertheless, the surplus-division interpretation is useful for providing intuition about Nash Bargaining.
Combining pairwise network stability and NiN bargaining, the base model gives us a solution concept that can predict equilibrium \((G^*, T^*)\) pairs. So, thus far, we have a model that in principle does endogenously capture network formation and bargaining (and in the next stage, premium setting). Nevertheless, the particular way that the base model accomplishes this job may be a source of concern, which I turn to next.

2.2.2 Problems with the Base Model

It is well documented in the literature (Gowrisankaran, Nevo and Town (2013); Lee and Fong (2013), etc.) that the implicit assumption in NiN bargaining (i.e., \(u_i\) and \(d_j\) taking the rest of \((G, T)\) as fixed when bargaining with one another) may be a source of concern. I show, using a stylized example, that this concern is more pronounced when endogenous network formation is involved. My example focuses on a case in which an insurer is bound by a network adequacy regulation to cover at least one of the two hospitals (absent the regulation, it would cover none\(^{14}\)). I show that the base model fails to capture competition from out of network hospitals and, hence, will highly over-predict how much the hospital, that the insurer will cover to abide by the mandate, will be able to leverage the mandate and charge the insurer.

**Setting of the example:** There are two hospitals \(u_1\) and \(u_2\) and one insurer \(d\). Bargaining parameters are symmetric: \(\gamma = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}\). The insurer’s premium is equal to 1\(^{15}\). If \(u_1\) is included, one unit of demand is generated. Thus the profit to \(u_1\) is \(t_1\) (assume no marginal cost) and that to \(d\) is \(1 - t_1 - 0.01\) where 0.01 is a fixed cost to \(d\) for including an additional hospital. Hospital \(u_2\) is a bit lower quality than \(u_1\) so the variable parts of both profits are multiplied by \(\vartheta < 1\). To keep the example simple, assume that the differentiation between \(u_1\) and \(u_2\) is only vertical. That is, if both are included, still one unit of demand is generated and all of that demand goes to \(u_1\). Figure \(3\) schematically presents the profits to all \(u_i\) and \(d\) as

---

\(^{14}\)This assumption is not realistic for a working insurance market. But it does capture the more general effect that I intend to highlight, while allowing for the setting to have only two hospitals, which significantly simplifies the discussion.

\(^{15}\)For simplicity, I do not assume here that the insurer sets its premium optimally. This assumption is not crucial but simplifies a lot.
functions of the \((G, T)\).^{16}

**Analysis of the example using the base model:** it is fairly straightforward to verify that the base model predicts the equilibrium market outcome: \((G^*, T^*) = \left( \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.495 \\ \emptyset \end{bmatrix} \right)\). Crucially, \(t^*_1\) is independent of \(\vartheta\).

That is, how good of a substitute \(u_2\) is for \(u_1\) is irrelevant in determining \(t^*_1\). Now consider a network adequacy regulation that imposes a large penalty of 100 on the insurer if it goes with an empty network. This will imply that the equilibrium will be \((G^*, T^*) = \left( \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 49.995 \\ \emptyset \end{bmatrix} \right)\). This means that \(u_1\) is extracting an extremely large surplus from \(d\) simply because the payoff to \(d\) would be -100 if it excluded \(u_1\). Again, the base model does not capture that instead of just excluding \(u_1\), the insurer \(d\) can replace it with \(u_2\). Next subsection introduces a model that addresses these issues.

### 2.2.3 Bargaining with threats of replacement

The objective of this subsection is to minimally expand on the bargaining conditions from the base model to deal with the issues illustrated above, while keeping the model computationally tractable. The basic intuition for the expansion is to allow the firm in danger of being substituted to anticipate this danger and bargain less aggressively. The definitions below formalize this idea.

**Definition 3** Under \((G, T)\), reimbursement rate \(t_{ij}\) is “safe for \(u_i\)” if \(\exists u_{i'}, t_{i'j}\) with \(g_{i'j} = 0\) such that the following hold:

---

\(^{16}\)I continue to suppress the notation on \(P\), as it has been assumed constant here.
\[ \pi_{d_j} \left( G_{ij}^{*-ij}, T_{ij}^{*-ij} + t_{ij}' \right) \geq \pi_{d_j} \left( G^*, T^* \right) \]  
\text{(5)}

\[ \pi_{u_{ij}'} \left( G_{ij}^{*-ij}, T_{ij}^{*-ij} + t_{ij}' \right) \geq \pi_{u_{ij}'} \left( G^*, T^* \right) \]  
\text{(6)}

\[ \pi_{u_{ij}'} \left( G_{ij}^{*-ij}, T_{ij}^{*-ij} \right) \geq \pi_{u_{ij}'} \left( G_{ij}^{*-ij}, T_{ij}^{*-ij} \right) \]  
\text{(7)}

where either profitability (condition 5) or incentive-compatibility (condition 6) holds strictly.

In words, reimbursement rate \( t_{ij} \) is “safe for \( u_i \)” if insurer \( d_j \) cannot strictly profit from replacing hospital \( u_i \) with hospital \( u_{ij}' \) in such a way that \( u_{ij}' \) is willing to participate. Conditions (5) through (7) formalize this idea. The profitability condition (5) says that \( d_j \) prefers replacing its current contract with \( u_i \) with a contract of \( t_{ij}' \) with \( u_{ij}' \) over the status quo. The incentive compatibility condition (6) says that \( u_{ij}' \) also prefers this move over the status quo. Finally, the no-commitment condition (7) says that after the substitution takes place, \( u_{ij}' \) would not prefer to drop its contract with \( d_j \). Trivial and unnecessary as it may seem, condition (7) plays an important role in ensuring the existence of an equilibrium in which multiple insurers are competing and enrollees can spill over among them.\(^{17}\)

**Definition 4** Under \((G,T)\), for a \( g_{ij} = 1 \), the “best safe reimbursement rate” for \( u_i \) is denoted \( \hat{t}_{u_i}(G,T,d_j) \) and defined as the highest profit to \( u_i \) among all values of \( t_{ij}' \) that would be safe for \( u_i \) if charged to \( d_j \).

Definition 4 provides us with the necessary notation for setting up the new bargaining conditions. The basic intuition is that at the equilibrium \((G^*, T^*)\), each non-null \( t_{ij}^* \) is equal to the Nash-bargaining reimbursement rate \( t_{ij}^{NB}(G,T,\gamma_{ij}) \) unless \( t_{ij}^{NB}(G,T,\gamma_{ij}) \) is not “safe” for \( u_i \), in which case, \( u_i \) will retreat to its best safe rate.

\(^{17}\)For more detail, see the online appendix.
Definition 5 \((G^*, T^*)\) satisfies “bargaining with threats of replacement” conditions if for \(\forall g^*_{ij} = 1\) we have:

\[
t^*_{ij} = \min\left( t^{NB}_{ij} (G, T, \gamma_{ij}), \hat{t}_{ui} (G, T, d_j) \right)
\] (8)

My new solution concept for \((G^*, T^*)\) combines pairwise stability conditions on network formation with bargaining-with-threats of replacement (rather than combining pairwise stability and Nash-in-Nash). Now let’s revisit the above example to see the implications of the new concept. Absent regulation, the equilibrium is \((G^*, T^*) = (\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} \min(0.495, 1 - \vartheta) \\ 0 \end{bmatrix})\). This implies that if \(u_2\) is a good substitute for \(u_1\), then \(u_1\) is forced to charge less to prevent a replacement by \(d\) with \(u_2\). Also under the regulation discussed in the example, the new equilibrium is: \((G^*, T^*) = (\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 - \vartheta \\ 0 \end{bmatrix})\), implying that an arbitrarily large penalty does not give \(u_1\) arbitrarily large leverage over \(d\).

In the online appendix, I provide non-cooperative foundations for formulation (8). I show that it is supported as Subgame Perfect Nash Equilibrium (SPNE) of the natural extension of the Rubinstein alternative offers game from a 1 \(\times\) 1 (i.e., one upstream firm and one downstream firm) to 2 \(\times\) 1. The proof in the appendix also provides economic intuition for some features of (8) that may seem \textit{ad hoc} in the first glance. For instance, it helps interpret the incentive compatibility (IC) condition in the definition of the highest safe price. Absent the non-cooperative foundation, the threat of replacement might be interpreted as the insurer making a take-it-or-leave-it offer to \(u_i\) upon disagreement with \(u_i\). Based on the non-cooperative game, however, a threat of replacement is a threat to cease negotiating with \(u_i\) and instead accept the competing offer by \(u_i\) which is already on the table for \(d_j\). That competing offer is expected to be the lowest \(u_i\) can afford to make, because \(u_i\) is currently excluded and is trying its best to replace \(u_i\) and enter \(d_j\)’s network. Another feature of (8) that the non-cooperative game explains is why \(\hat{t}_{ui} (G, T, d_j)\) does not impact \(t^*_{ij}\) if it is higher than \(t^{NB}_{ij} (G, T, \gamma_{ij})\) (i.e., why the minimum.) I leave the details of the intuition behind this feature to the appendix. There, I also discuss why chose the model in this paper over other alternatives such as cooperative-based models, non-cooperative models such as Abreu and Manea (2012), vertical contracting models such as Segal (1999); Rey and Whinston (2013); Prat and Rustichini (2003), and
bargaining models such as Stole and Zwiebel (1996a, b) among others. The rest of the paper is about the application.

3 Background

This section (i) describes the nature of price consequences that I study for network-adequacy regulations; and then (ii) introduces “CommCare”, the market that I study.

3.1 Narrow Hospital-Networks and Price-Consequences of Regulating Them

There are multiple reasons why an insurer would prefer to form a narrow hospital network. Among those mechanisms, the following are most frequently noted: (i) The insurer might exclude expensive hospitals so that it can charge lower premiums and attract more enrollees; (ii) it might exclude expensive (and high-quality) hospitals in order to attract only healthier enrollees; (iii) the insurer might exclude some hospitals in order to steer patients to a select set of other hospitals. In exchange for the additional volume directed to them, the included hospitals offer the insurer lower reimbursement rates than they would absent the exclusion of their rivals. All of these provide cost-saving opportunities for the insurer, which will be mitigated if narrow networks cannot be formed.

But none of the above is the main focus of this paper. What the above mechanisms have in common is that they all lead to narrow networks that exclude more expensive hospitals and include cheaper ones. In contrast, the narrowest-network plan in CommCare includes the two most expensive hospitals in Boston, but excludes cheaper hospitals that are geographically close to those two. In section (i), I will argue that this results from lack of economies of scale. Thus, in this paper, I will not focus on adverse selection.

18Mechanisms (i) and (ii) are by construction about forming narrow networks comprised of cheaper hospitals. To see why mechanism (iii) does the same, note that the high-volume-in-exchange-for-low-rate nature of it can be thought of as a volume-based (i.e., non-linear) contract implemented relationally. The most aggressive volume-based contract that hospital $u_i$ can offer would be a two-part tariff with a per-unit price equal to $u_i$’s marginal cost $c_i$. Even this contract cannot incentivize the insurer to exclude any hospital $u_i$ that has a lower marginal cost than $u_i$. 

16
(i.e., I assume perfect risk-adjustment), or volume-based contracts although they could be easily incorporated into the model. Mechanism (i) is naturally incorporated into the model though in my application it is not what generates the narrow network.

The type of price consequences that this paper studies is more general. What I study is the extent to which a network-adequacy regulation is taken advantage of, in the bargaining process, by hospitals because they know that the insurer has a harder time credibly threatening to exclude them. Unlike the mechanisms above, these price consequences arise from the compromise of the insurer’s threats of exclusion rather than exclusion itself. To illustrate this, if a plan’s network size is $\%X$, a regulation that mandates a minimum of $\%X$ may still raise the prices for that plan without changing its network. This direct price effect has been mentioned as a major source of concern about network adequacy regulations in the literature (e.g., [Howard (2014)]). It can be present in any market regardless of the reason why the networks are narrow in that market. Therefore, this paper’s insights are generalizable to settings where narrow networks arise for a different reason than they do in CommCare.

Before moving to estimation, I discuss an important possible criticism to the significance of the type of price consequences that this paper focuses on. It may be argued that the insurer can avoid the compromise of its bargaining position by forming a hospital network that is strictly broader than the minimum mandated size (e.g., when the minimum size is $\%50$, the insurer can include $\%50$ plus at least one more hospital system.) This way, no hospital will be pivotal anymore to the insurer’s ability to abide by the regulation.

The criticism above assumes that the insurer is foresighted: it anticipates the consequences of being at the boundary of the regulation; and, hence, goes strictly above it. However, if we also assume a similar foresight for the hospitals, they should know the intent of this move by the insurer. Each hospital now knows it is pivotal to the insurer’s ability to escape the boundary of the regulation; and will charge the insurer for it. Thus, the same bargaining-position shift re-emerges, rendering the inclusion of the extra hospital ineffective and unnecessary. Therefore, the price consequences that I study are present with or without such foresight. Hence I abstract away from this foresight, which allows me to focus on the consequences of the regulation, while helping to keep this already complex model as parsimonious is possible.
3.2 The CommCare Market

CommCare –or, more precisely, the “Commonwealth Care” market– was a subsidized health insurance exchange where people with low income (below 300% of the federal poverty line) who could not get insured by an employer or public programs were eligible to enroll. CommCare was operated by the state of Massachusetts as part of the Massachusetts Healthcare Reform (which preceded the ACA.) I chose CommCare for analysis because I believe that it is the most similar market to the ACA exchanges among those health insurance markets for which comprehensive data on claims, plan enrollments, network structures, and premiums are available. Both markets are subsidized health insurance exchanges for individuals in which private insurers compete by offering highly standardized plans. Nevertheless there are some differences between the two markets. In section 5 I will discuss how some of these differences might lead us to expect different responses to network adequacy regulations between CommCare and the ACA exchanges.

CommCare was established in 2007 and had around 200,000 enrollees annually in fiscal years of 2011 through 2013. Enrollees were partitioned into three income-groups: below the Federal Poverty Line (FPL), between 100% and 200% of FPL, and between 200% and 300% of FPL. Each income group was subject to a different subsidy rate and would pay a different (uniform) co-pay when visiting an in-network hospital. In particular, the below-FPL

---

19 Other common features between CommCare and ACA exchanges include but are not limited to: (1) Mandatory participation for people who are eligible and cannot find insurance elsewhere; (2) Risk-adjustment programs to discourage insurers from competing for healthier enrollees.

20 Some of the differences are: (1) CommCare included below poverty people but ACA exchanges leave that group for state Medicaid programs. (2) ACA exchanges also include people ineligible for subsidies but CommCare left them for another program named CommChoice, short for “Commonwealth Choice”. (3) CommCare was more highly standardized than ACA exchanges are. In the ACA exchanges, plans can be offered in different “metal tiers” with different benefits. In CommCare, there was no deductible or co-insurance. There was only a small co-pay which was not to be decided by the insurer. It was the same for each income group across all insurers and all hospitals. (MassHealth, 2011). (4) The subsidies for each income group are fixed in the ACA exchanges but were linked to premiums in CommCare, meaning in CommCare most consumers would receive more in subsidies if they bought a higher-premium plan.

21 In fact, more precisely, the total number of income group was five rather than three: below FPL, 100%-150% of FPL, 150%-200% of FPL, 200%-250% of FPL, and 250%-300% of FPL. Unfortunately, in my data, I cannot distinguish between 100%-150% of FPL and 150%-200% of FPL or between 200%-250% of FPL and 250%-300% of FPL. Therefore, I
group was fully subsidized and paid no premium. This group also paid zero co-pay. Other groups were also heavily subsidized but paid positive premiums and co-payments. Given that the full subsidization of below-poverty enrollees rendered them price insensitive, CommCare imposed a new, auction-like, regulation from 2012 on in order to foster price competition among insurers. The regulation roughly stated that below-poverty consumers who had just joined the market were only allowed to choose between the two cheapest plans.

I examine the whole CommCare market for estimation and identification. But for the counterfactual policy I apply some restrictions. First, I focus on the 2011 fiscal year. I do this because the particular auction-like regulation for 2012 and 2013 complicates the premium-setting game. Also, to save on computation, I restrict my analysis geographically by concentrating on the Greater Boston Area (which I define by all zip codes not more than 30 miles from zip code 02114 in downtown Boston). Finally, I focus on general acute care hospitals rather than all hospitals (throughout this paper, unless otherwise stated, by “hospital” I mean a general acute care hospital.) There are 28 hospitals in the Greater Boston Area, which are owned by 16 hospital systems. Figure 4 locates these hospitals on a map.

Figure 4: General Acute Care Hospitals in Massachusetts. Those in the Greater Boston Area are shown in blue rectangles.
In 2011 (and thereafter, until 2014 when CommCare was shut down and the ACA took over), five insurers competed in CommCare. One of them (Fallon) was inactive in the Boston area. I focus, therefore, on the other four: BMC (Boston Medical Center), Celticare, NHP (Neighborhood Health Plan), and Network Health. Given that below poverty consumers in CommCare are insensitive to premiums, CommCare sets a cap on premiums in order to prevent arbitrary price increases. In 2011, the cap was $425/month per person. CommCare also had a mandated premium floor (likely to help prevent adverse selection by shifting the competition from the price domain to quality domain) which was at the level of $404/month in 2011. Table 1 summarizes the state of the market in the 2011 fiscal year. Celticare has the narrowest network among the four with only 4 systems covered out of the whole 16. It also charges the lowest premium. Nevertheless, my estimations indicate that Celticare pays, on average, a higher reimbursement to its in-network hospitals for a severity-adjusted admission than any of the other three insurers. The main reason is that Celticare covers the two most expensive hospitals in Massachusetts: Massachusetts General Hospital (MGH) and Brigham and Women’s Hospital (BWH). Both MGH and BWH belong to the Partners Healthcare system (“Partners” henceforth). MGH and BWH are highly prestigious academic medical centers and attract a lot of patients in any network in which they are participate, and particularly if the network is a smaller one like Celticare’s, leading to high average reimbursements as well as high average marginal cost of inpatient care. All of the other plans have broader networks, higher premiums, and lower estimated average reimbursements than Celticare.

Celticare has other major differences from the other insurers in the market. Celticare has a substantially lower market share than the rest of the plans. Also, unlike BMC, NHP, and Network Health, which are all based in Massachusetts and entered CommCare in fiscal year of 2007, Celticare is based in St. Louis, Missouri, and entered the market in 2010. In other words, Celticare is a smaller insurer compared to the other three. As section 4 will explain in more detail, Celticare’s small enrollment size and little-known brand indeed have implications for its hospital network and premium. There is evidence from the ACA exchanges that smaller insurers have a harder time building hospital networks due in part to lack of established relationships with agents, as well as lack of sufficient number of enrollees to diversify risk (McKinsey (2015)). Celticare itself cites its “small size” as its biggest limitation to engage with providers (Health Policy Commission, 2013). Also,
Table 1: The CommCare Market in FY 2011, Greater Boston Area

<table>
<thead>
<tr>
<th></th>
<th>BMC</th>
<th>Celticare</th>
<th>NHP</th>
<th>Network Health</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systems Covered (out of 20)</td>
<td>11</td>
<td>4</td>
<td>15</td>
<td>14</td>
</tr>
<tr>
<td>Premium ($/month)</td>
<td>425</td>
<td>404</td>
<td>425</td>
<td>425</td>
</tr>
<tr>
<td>Average reimbursement ($/admission)</td>
<td>11,781</td>
<td>14,120</td>
<td>12,025</td>
<td>12,069</td>
</tr>
<tr>
<td>Average inpatient marginal cost ($/admission)</td>
<td>6721</td>
<td>10302</td>
<td>8335</td>
<td>8362</td>
</tr>
<tr>
<td>Market Share</td>
<td>32%</td>
<td>12%</td>
<td>27%</td>
<td>29%</td>
</tr>
<tr>
<td>Based in</td>
<td>MA</td>
<td>MO</td>
<td>MA</td>
<td>MA</td>
</tr>
<tr>
<td>First Fiscal Year in CommCare</td>
<td>2007</td>
<td>2010</td>
<td>2007</td>
<td>2007</td>
</tr>
</tbody>
</table>

according to McKinsey (2015), less well-known insurers like Celticare “may face greater pressure to be price competitive to attract members” to make up for their weak brand names. Section 4 shows how different features of my formulations for profit functions $\pi_{ui}(G, T, P)$ and $\pi_{dj}(G, T, P)$ capture these differences among plans (in particular Celticare’s low economies of scale) as well as the implications of these differences for network formation, bargaining, and premiums.

## 4 Estimation

The model’s purpose was to predict the market outcome $(G^*, T^*, P^*)$ given all profit functions $\pi_{ui}(\cdot)$ and $\pi_{dj}(\cdot)$ and bargaining parameters matrix $\gamma_{m \times n}$. Estimation is the reverse. We observe $(G^*, T^*, P^*)$ plus some partial data about $\pi_{ui}(\cdot)$ and $\pi_{dj}(\cdot)$, and our objective is fully backing out estimates $\hat{\pi}_{ui}(\cdot)$, $\hat{\pi}_{dj}(\cdot)$, and $\hat{\gamma}_{ij}$ for all $i, j$ so that we can then do counterfactual analysis. Of course, part of the identification comes from parametric assumptions on profit functions and bargaining powers. I start by specifying how profit functions depend on demand and cost functions. Later (in sections 4.2 through 4.5), I will further specify the assumptions made on demand functions, cost functions, and the bargaining parameter matrix.

Hospital $u_i$’s profit from each insurer $d_j$ that covers $u_i$ is how many units of healthcare per month care it provides to enrollees of $d_j$ times the marginal profit $u_i$ makes from each unit of care provided to $d_j$ enrollees. Different

---

23 One “unit of healthcare” is one hospital admission with average severity. Severities are measured in DRG weights and discussed in section 4.2.
consumer groups $\kappa$ may have different tastes for hospitals and insurers (I will discuss in more detail how consumers are binned into different $\kappa$ groups in section 4.2). So, $u_i$’s total profit is given by:

$$\pi_{u_i}(G, T, P) = \sum_{\kappa} \sum_j (D_j^{\kappa}(G, P) \times \sigma_{ij}^{\kappa}(G) \times (t_{ij} - c_i))$$  \hspace{1cm} (9)$$

In (9), $c_i$ is the marginal cost of providing one unit of care to a patient. Function $D_j^{\kappa}(G, P)$ depicts how many enrollees from bin $\kappa$ choose to enroll with insurer $d_j$. Function $\sigma_{ij}^{\kappa}(G)$ represents how many units of care per month the average $d_j$ enrollee from bin $\kappa$ receives from hospital $u_i$.24,25

Insurer $d_j$’s profit is the sum of premiums it charges minus the different types of costs it incurs:

$$\pi_{d_j}(G, T, P) = \sum_{\kappa} \left( D_j^{\kappa}(G, P) \times \left( p_j - \xi_j - \sum_i (\sigma_{ij}^{\kappa}(G) \times t_{ij}) \right) \right) - \sum_{i \text{ s.t. } g_{ij} = 1} f_{ij}$$ \hspace{1cm} (10)

where $p_j$ is the premium that $d_j$ charges. The first element of cost to $d_j$ is $\xi_j$, which denotes monthly “non-inpatient costs” to $d_j$ per each enrollee. $\xi_j$ can include costs from out-patient care by hospitals, pharmacies, non-hospital care, the variable component of administrative costs, etc. The second element is the amount that $d_j$ pays to each in-network hospital $u_i$ in reimbursements, which is equal to $\sigma_{ij}^{\kappa}(G) \times t_{ij}$ for an average enrollee of type $\kappa$.

The third component of costs to $d_j$ is the set of fixed costs $f_{ij}$ which $d_j$ incurs for every hospital $u_i$ that it covers. These fixed costs are an important part of what explains the differences in economies of scale across plans (the other key part is heterogeneity in plans’ brand values, and is discussed in section 4.3). They explain why Celticare forms a narrow network including prestigious hospitals like MGH and BWH but excluding a lot of cheap hospitals geographically close to them.26 There are multiple sources

---

24 Here, neither $D_j^{\kappa}(G, P)$ nor $\sigma_{ij}^{\kappa}(G)$ is a function of $T$. The reason is that in CommCare, there was no co-insurance. Where there is co-insurance, $T$ needs to be included as an argument to both functions.

25 As I will formally define soon, by construction, $\sigma_{ij}^{\kappa}(G) = 0$ for all $ij$ such that $g_{ij} = 0$.

26 For instance, in the absence of such fixed costs, one would not expect low-cost hospitals to be left out of hospital-networks that cover high-cost hospitals; because covering a hospital that would agree to join the network with a low price could steer patients away from more expensive hospitals, and, in the worst case of steering no one away, would do no harm to the insurer.
for such fixed costs. First, the bargaining process itself is costly. Second, “risk-diversification” can be modeled as insurers facing fixed costs. Third, insurers sometimes leave hospitals out of the network if they fail to meet some quality standards that may not be directly observable to patients. I model such quality concerns as part of the \( f_{ij} \) costs. Finally, the fourth source is insurer \( d_j \)'s over- or under-estimation of how profitable signing a contract with hospital \( u_i \) would be. This latter term plays the role of the structural errors that will be used in the estimation procedure to rationalize the data.

Given the specifications of \( \pi_{u_i}(\cdot) \) and \( \pi_{d_j}(\cdot) \) in (9) and (10), estimating the profit functions means estimating all \( \hat{\sigma}_{ij}(\cdot), \hat{D}_{kj}(\cdot), \hat{\xi}_j, \) and \( f_{ij} \) for all \( i \) and \( j \). The procedure that I develop for this estimation has four steps which correspond to the four steps of the sequential game (shown in figure 2) in reverse order. Figure 5 schematically represents the four steps of the estimation process. The first two steps are based heavily on methods developed by Capps, Dranove and Satterthwaite (2003) and Ho (2006). In step 1, I estimate all \( \hat{\sigma}_{ij}(\cdot) \) using a logit model of hospital choice. In step 2, given the outcome of step 1, I estimate all \( \hat{D}_{kj}(\cdot) \) using a model of plan choice. In step 3, given the full demand estimation outcome from steps 1 and 2, I impose the Nash Bertrand assumption on premium-setting to back out all non-inpatient costs \( \hat{\xi}_j \). Finally, in step 4, I impose the network-formation and bargaining model developed in section 2 to back out all \( \hat{f}_{ij} \) and \( \hat{\gamma}_{ij} \). In the remainder of this section, I review the data and discuss each of the four steps of the estimation procedure in more details. For each step, after further parameterization of

---

27 Insurers with smaller sizes (e.g., Celticare in CommCare) often have a harder time contracting with hospitals since the expected numbers of their enrollees who would visit a hospital are less likely to be large enough to make the insurer feel more confident that those enrollees, who turn out to be more costly to the insurer, will be cancelled out by those who are less costly. Of course, the most accurate way to model this risk-diversification issue would be to (1) directly model the financial risk that insurers face regarding the healthcare costs of their enrollees, and (2) model insurers as risk-averse agents. This would substantially complicate the model. However, modeling this phenomenon as a fixed cost does not further complicate the model and, at the same time, captures the idea that a small insurer might have a harder time sign a contract with a hospital without facing risks.

28 For instance, if a hospital has old equipment which the insurer believes might have serious side effects for patients, the insurer may decide to leave the hospital out of network. Whether this comes from altruism towards customers or from concerns about future profits, it can be captured as a cost to the insurer for including that hospital in its network.
Figure 5: The four steps of the estimation procedure

the object of interest if necessary, I (i) describe the estimation methodology, (ii) discuss the identification, and (iii) report and interpret the results.

4.1 Data

Datasets used in my analysis consist of (1) Data on hospital discharges and medical claims, (2) data on insurance plan enrollments, and (3) data on hospital and insurance plan characteristics. Below, I discuss these different data in more details.

4.1.1 Data on Hospital Discharges and Medical Claims

I use data on hospital discharges and medical claims at two points. First, I use data on payments from insurers to hospitals in order to construct a measure of the reimbursements matrix $T^*$, the details of which are explained in the online appendix. Second, I use data on hospital discharges in step 1 of the estimation procedure (see figure 5) where I back out hospital choice functions $\hat{\sigma}_{ij}(\cdot)$. The primary source of my data for discharges and medical claims is the Massachusetts All Payers Claims Database (MA-APC) from the Center for Health Information and Analysis (CHIA). MA-APC has a medical claims dataset which offers very rich and comprehensive information on medical claims and discharges from 2010 to 2014.

In the medical claims dataset, the unit of observation is claim-line, which pertains to an individual medical bill that a provider (e.g., physician, hospital, pharmacy) sends to an insurer. For each claim-line, MA-APC’s medi-
cal claims dataset contains information on patient demographics, diagnosis, type of claim (in particular whether or not the claim pertains to an inpatient hospital admission), date, payments, and identifiers for the provider, the patient, and the insurer. The demographic information consists of the patient’s gender, age, and the 5-digit zip code of residence. The diagnosis information is reported in ICD-9-CM diagnosis codes. The payments are broken down into payments by the patient and payment by the carrier (i.e., insurer). The provider ID is the National Provider Identifier (NPI) issued to providers by the Centers for Medicare and Medicaid Services (CMS). A hospital may have multiple NPIs. The patient identifier is an ID assigned by CHIA. This is an “APCD internal ID” which enables researchers to link together different claims and enrollment records of the same patient within the APCD.

I carry out several further processings on the medical claims data before I use it in the estimation procedure. I restrict the medical claims dataset to inpatient claims only. I also use data on the NPIs of general acute care hospitals in Massachusetts from CMS both to restrict the claims data to those from general hospitals in Massachusetts and to match each claim line in the data to a hospital name. Since a hospital may issue multiple bills to an insurer for different services provided during a single hospital stay, I aggregate the claims data from the “claim-line level” to the “admission-episode” level by lumping together all claims that have the same patient ID, insurer ID, hospital name (I do not use hospital NPI, since a hospital can have multiple NPIs), diagnosis, and service provision year and month. Finally, I link the ICD-9-CM diagnosis codes to a coarser category called Clinical Classifications Software (CCS) developed by the Agency of Healthcare Research and Quality (AHRQ). This linkage enables me to link diagnoses to MS-DRG severity indices for different diagnosis groups.

4.1.2 Data on Insurance Plan Enrollments

I use data on enrollments in CommCare insurance plans to back out plan demand functions $\hat{D}_{j}(\cdot)$ from consumers’ plan choice patterns in step 2 of

---

29ICD-CM is short for International Classification of Diseases, Clinical Modification. The ICD-9-CM is an adaptation by the U.S. National Center for Health Statistics of ICD-9 diagnostic codes which are used internationally for diagnosis classification. There are about 14,000 ICD-9 codes.

30This ID cannot be used to personally identify any patient or enrollee.
the estimation procedure (see figure 5 and section 4.3). My enrollments data comes from the MA-APCD’s enrollments dataset.

In the enrollments dataset, the unit of observation is enrollment record. An enrollment record is uniquely identified by its enrollee ID and fiscal year. Enrollee IDs are the same IDs as those assigned to patients in the medical claims file so the two files can be linked together. Fiscal year is a variable that I construct for each record from enrollment start year and month, as well as publicly available information on the timing of CommCare’s operation. A CommCare fiscal year started on July with an open enrollment period, making, for instance, July of 2010 the first month of the 2011 fiscal year (henceforth, 2011FY). For each enrollment record, MA-APCD’s enrollments dataset contains information on enrollee demographics and insurer ID. Demographic information in the enrollments dataset consists of the same demographic elements that were included in the medical claims file plus an additional one: the enrollee’s income group, which can be one of the following three groups: below FPL, within 100%-200% of FPL, and within 200%-300% of FPL.

4.1.3 Data on Hospital and Insurance Plan Characteristics

I supplement the MA-APCD medical claims data and enrollments data with several other datasets. I use public data from the Mass Connector on CommCare plans’ hospital networks. I use this dataset in the hospital choice estimation (see step 1 in figure 5) to identify the choice set of each patient who visits a hospital. I also use it in the plan demand estimation (see step 2 in figure 5), in conjunction with the results of the hospital choice estimation, in order to construct measures of relative values of different plans’ hospital networks to different consumer bins $\kappa$ on CommCare. I also use data from the Mass Connector on CommCare plans’ premiums as well as subsidy rates in CommCare. I use these data in the plan demand estimation process (see step 2 in figure 5).

Finally, I use data on hospitals’ costs from CMS’s Healthcare Cost Report Information System (HCRIS). These data include annual costs reports by all Medicare-certified providers (which include all of the hospitals I study) to CMS, broken down into seven cost centers –such as general service, inpatient

\footnote{Even though data on hospital networks, plan premiums, and subsidies in CommCare was public, it was removed from the Mass Connector website before I could access it. I thank Mark Shepard for sharing these data with me.}
service, outpatient service, and ancillary service— and then, for each cost center, broken down in great details into many items. I then adopt the approach used in Schmitt (2015) to construct from these cost items a measure of $c_i$, the marginal inpatient costs per severity adjusted hospital admission. For details on how I construct these $c_i$ measures, see the online appendix.\(^{32}\)

### 4.2 Step 1: Estimating Hospital Demand

In the first step of estimating the model, I back out the hospital choice function $\sigma_{ij}^\kappa(G)$. This function represents the total demand for hospital $u_i$ from an average consumer who is in bin $\kappa$ and has enrolled in insurance plan $d_j$. This is a weighted sum of demand levels $\sigma_{ij}^{\kappa,\psi}(G)$ for different diagnoses $\psi$, where the weight for each diagnosis is its DRG severity measure $w_\psi$. Formally:

$$\sigma_{ij}^\kappa(G) = \sum_{\psi \in \Psi} w_\psi \times \sigma_{ij}^{\kappa,\psi}(G) \quad (11)$$

Equation (11) has a direct implication for the interpretation of (9) and (10). It implies that the reimbursement made by insurer $d_j$ to hospital $u_i$ for a hospital admission with diagnosis $\psi$ is equal to $w_\psi \times t_{ij}$. Similarly, the marginal cost to hospital $u_i$ of providing care to an admitted patient with diagnosis $\psi$ is implicitly assumed equal to $w_\psi \times c_i$. This linearity assumption is standard in the literature (see, for example, Gowrisankaran, Nevo and Town (2013); Ho and Lee (2017b); Prager (2015)).

Diagnosis-specific hospital choice function $\sigma_{ij}^{\kappa,\psi}(G)$ is assumed to come from a multi-nomial logit model of hospital choice for different consumer bins $\kappa$. I bin the consumers based on two observables: the 5-digit zip code of residence location, and income group (below poverty, between 100% and 200% of poverty, and between 200% and 300% of poverty). Therefore, bin $\kappa$ is a combination of location $l$ and income group $y$. For each individual $k$, I use the notation $\kappa(k)$ (notation picked up from Ho and Lee (2017b)) to represent the consumer bin that consumer $k$ belongs to. Notations $l(k)$ and $y(k)$ are also used in a similar manner.

---

\(^{32}\)The original cost reports data from HCRIS is very raw and requires a great deal of processing before $c_i$ measures can be constructed off of it. I thank Matt Schmitt for sharing with me a cleaned up version of the HCRIS data.
Underlying my multi-nomial choice model is the following utility function for individual \( k \) with diagnosis \( \psi \) admitted to hospital \( u_i \):

\[
V_{ik\psi}^H = \delta d_{il(k)} + v_i \times \left(1 + \theta^{\text{diag}} \times w_\psi\right) + \varepsilon_{ik\psi}
\] (12)

This utility function assumes that when evaluating hospital \( u_i \), consumer \( k \) pays attention to how far the hospital is located from where she lives as well as to the quality of the hospital. In (12), \( d_{ik} \) is the distance between hospital \( i \) and the residence of individual \( k \) who belongs to bin \( \kappa(k) \), and \( \delta \) is the corresponding coefficient. Also \( v_i \) captures hospital \( i \)'s quality as perceived by consumers. The utility function in (12) allows for the possibility that for more severe conditions, consumers care more about hospital quality and less about distance. Therefore, in (12), there is also an interaction term \( v_i \times w_\psi \) is multiplied by the coefficient \( \theta^{\text{diag}} \), which measures the extent to which patients with more severe conditions (i.e., higher \( w_\psi \)) pay extra attention to hospital qualities \( v_i \) compared to patients with less severe conditions. The last term in (12), \( \varepsilon_{ik\psi} \) is an idiosyncratic error term whose distribution is i.i.d Type 1 extreme value with variance of 1. The the distribution of \( \varepsilon_{ik\psi} \) gives a closed-form representation for \( \sigma_{ij}^{\kappa,\psi}(G) \). For \( g_{ij} = 0 \), we have \( \sigma_{ij}^{\kappa,\psi}(G) = 0 \). For \( g_{ij} = 1 \) we get:

\[
\sigma_{ij}^{\kappa,\psi}(G) = \lambda_{\kappa,\psi} \frac{e^{V_{ik\psi}^H}}{\sum_{l' \text{ s.t. } g_{ij}=1} e^{V_{l'\kappa\psi}^H}}
\] (13)

where \( \bar{V}_{ik\psi}^H \) (note that in the subscript, it is \( \kappa \) and not \( k \)) the average of \( V_{ik\psi}^H \) over all consumers \( k \) of type \( \kappa \). That is, \( \bar{V}_{ik\psi}^H = \delta d_{il} + v_i \times \left(1 + \theta^{\text{diag}} \times w_\psi\right) \), where \( l \) is the location element of \( \kappa \). Also \( \lambda_{\kappa,\psi} \) is the expected rate of hospital admission with diagnosis \( \psi \) per month per member for consumer bin \( \kappa \).

**Estimation Procedure:** The parameters to estimate in this step are \( (\hat{\delta}, \hat{\theta}^{\text{diag}}, \hat{\lambda}_{\kappa,\psi}) \). I observe \( \hat{\lambda}_{\kappa,\psi} \) directly from the data. To estimate the rest of the parameters, I estimate a multi-nomial logit model. I observe hospital choices \( \sigma_{ij}^{\kappa,\psi}(G) \) for all \( i, j, \kappa, \psi \) in my data. I estimate the parameters using a maximum likelihood approach, matching the observed \( \sigma_{ij}^{\kappa,\psi}(G) \).

**Identification:** Hospital qualities \( v_i \) are identified by cross-hospital variation in discharge volumes. If hospital \( u_i \) has a higher discharge share than

\[33\] I construct \( \sigma_{ij}^{\kappa,\psi}(G) \) for each \( i, j, \kappa, \psi \) using the discharge data for choices made by patients and data on the network \( G \) for the choice set.
hospital $u_i'$ among consumer bins $\kappa$ that live in locations equally far from $u_i$ and $u_i'$, then $v_i$ must be larger than $v_i'$. Distance coefficient $\delta$ is identified by within-hospital, cross-location variation in discharge volumes. The faster the discharge share of hospital $u_i$ diminishes as we look at consumer bins $\kappa$ farther from $u_i$, the more negative the distance coefficient $\delta$ must have been. The severity weight coefficient $\theta_{diag}$ is identified by the variation in discharge volumes within consumer bins $\kappa$ but across hospitals and diagnoses. The more the discharge shares get skewed towards higher quality hospitals as we look at more severe diagnoses $\psi$, the higher the implied $\theta_{diag}$ would be.

**Results and Interpretation:** Table 2 summarizes the results of the multi-nomial logit estimation of the hospital choice model. The distance coefficient is estimated to be $\hat{\delta} = -0.137$, which is consistent with the literature. It implies that an extra 10 miles of distance reduces the share of a hospital by an average of 29%. The DRG weight coefficient is estimated to be $\hat{\theta}_{diag} = 0.036$. To illustrate the interpretation of this number, suppose hospitals $u_i$ and $u_i'$ are such that $\frac{\hat{v}_i - \hat{v}_{i'}}{\hat{\delta} \times (1 + \hat{\theta}_{diag})} = 10$. That is, the average patient with diagnosis $\psi$ of severity $w_{\psi} = 1$ would choose $u_i'$ over $u_i$ if $u_i'$ is no more than 10 miles farther from her than $u_i$. Then, a DRG coefficient of $\hat{\theta}_{diag} = 0.036$ would imply that the average patient with diagnosis $\psi'$ of severity $w_{\psi'} = 2$ would be willing to travel $10 \times \frac{1 + 2 \times 0.036}{1 + 0.036} \approx 10.36$ more miles to visit hospital $u_i'$ than she would to visit $u_i$.

**Hospital Choice Estimates**

<table>
<thead>
<tr>
<th></th>
<th>coeff.</th>
<th>std. error</th>
</tr>
</thead>
<tbody>
<tr>
<td>distance (miles)</td>
<td>-0.136***</td>
<td>(0.002)</td>
</tr>
<tr>
<td>DRG weight</td>
<td>0.036**</td>
<td>(0.017)</td>
</tr>
<tr>
<td>Hospital FEs</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Num. hospital admissions</td>
<td>40,247</td>
<td></td>
</tr>
<tr>
<td>pseudo $R^2$</td>
<td>0.413</td>
<td></td>
</tr>
</tbody>
</table>

std errors in parentheses, *: $p < 0.1$, **: $p < 0.05$, ***: $p < 0.01$

Table 2: Hospital Choice Model
4.3 Step 2: Estimating Insurance Plan Demand

Having estimated a model of hospital choice, I now turn to estimating the insurance plan demand functions \( D_j^\kappa(\cdot) \). I assume that consumer \( k \)'s valuation of insurance plan \( j \) comes from the following utility function:

\[
V_{jk}^I = \alpha \times EU_{j\kappa(k)} + \beta_{y(k)} p_{jk}^{sub} + \Delta_j + \epsilon_{jk} \tag{14}
\]

where \( EU_{j\kappa(k)} \) is the expected utility of insurer \( d_j \)'s network of providers for consumer \( k \) of type \( \kappa \), and \( \alpha \) represents how much consumers care about network expected utility when they compare plans. Network-of-providers utility \( EU_{j\kappa} \) is given by:

\[
EU_{j\kappa} = \sum_{\psi} \lambda_{\kappa,\psi} EU_{\psi jk} \sum_{\psi} \lambda_{\kappa,\psi} \tag{15}
\]

which means that \( EU_{j\kappa} \) comes from a weighted average of \( EU_{\psi jk} \), which are values of plan \( j \)'s network of providers for different diagnoses \( \psi \) for a consumer of type \( \kappa \). The weights come from the likelihoods of different conditions for type \( \kappa \). Given the distribution of \( \varepsilon_{ik\psi} \), there is a closed-form representation for how the values \( V_{i\kappa\psi}^H \) of hospitals covered by \( d_j \) contribute to \( EU_{jk}^\psi \):

\[
EU_{jk}^\psi = \ln \left( \sum_{i \text{ s.t. } g_{ij} = 1} e^{\frac{V_{ijk}^H}{\lambda_{\kappa,\psi}}} \right) \tag{16}
\]

\( \Delta_j \) in \( 14 \) is insurer \( d_j \)'s fixed effect, which can be interpreted as “brand effect”. \( \beta_{y(k)} \) is the sensitivity of consumer \( k \), who belongs to income group \( y(k) \), to the monthly premium \( p_{jk}^{sub} \) she has to pay for plan \( j \). Due to subsidization in CommCare, there is a difference between the premium \( p_j \) each insurer charges for its plan and \( p_{jk}^{sub} \), the one a consumer buying that plan pays. The relationship between \( p_{jk}^{sub} \) and \( p_j \) is linear. There is a fixed subsidy \( b_y \) and a pass-through rate \( a_y \) for each income group forming a linear
relationship, as shown in (17). All $a_y$ and $b_y$ are observed.

$$p_{jk}^{sub} = a_{y(k)} \times p_j - b_{y(k)} \quad (17)$$

Finally, in (14), $\epsilon_{jk}$ is an i.i.d extreme value type I error term, which gives us the following closed-form solution for $D_\kappa^*(G, P)$:

$$D_\kappa^*(G, P) = \Lambda_\kappa \frac{e^{\bar{V}_I^\kappa}}{\sum_{j'} e^{\bar{V}_{I'}^\kappa}} \quad (18)$$

where (similar to the construction of $\bar{V}_{H_{i\psi}}$) $\bar{V}_{I\kappa}$ is constructed by averaging $V_{jk}^I$ over all consumers $k$ of type $\kappa$ (which gives the same formula as $V_{jk}^I$ but without the error term $\epsilon_{jk}$). Also, $\Lambda_\kappa$ is the population of individual consumers of type $\kappa$.

**Estimation Procedure:** The parameters to estimate in this step are coefficients on network utility and monthly premium, as well as the brand effects: $(\hat{\alpha}, \hat{\beta}_y, \hat{\Delta}_j)$. Populations $\Lambda_\kappa$ are observed, and network utilities $EU_{jk}$ can be constructed from the results of the hospital choice estimation (and are hence, treated as data in this step). To back out $(\hat{\alpha}, \hat{\beta}_y, \hat{\Delta}_j)$, I estimate the multi-nominal logit model set up in equations (14) through (18) using a maximum likelihood approach.

**Identification:** Brand fixed effects $\Delta_j$ are identified by the variation in the enrollment volumes across plans. Network utility coefficient $\alpha$ is identified by within-plan, cross-location variation in enrollment volumes. To illustrate, if the ratio of Celticare enrollments to NHP enrollments for a certain income group in a certain year is constant across residents of different zip codes, then all of the difference is picked up by brand effects. But if this enrollments ratio is lower in locations where NHP’s hospital network includes some nearby hospitals, but Celticare’s does not, then this variation in ratios is explained by

---

34 There is one non-linearity in CommCare’s subsidization policy. For the second poorest income group (i.e., between 100% and 150% of poverty), $b_y$ is determined in such a way that there is at least one free option for income group $y$. That is, $b_y$ should equate $\min_j p_{jk}^{sub}$ to zero for that group. In my counterfactual analysis, I do not incorporate this effect. Instead, I take all $a_y$ and $b_y$ remains constant in response to counterfactual scenarios. This simplifies the model greatly and does not substantially change the results.

35 In 2011, an X% of the premiums charged by insurers in CommCare was subsidized through fixed subsidies (i.e., $a_y$) and another Y% through variable subsidies (i.e., $b_y$). Consumers paid the remaining 100-X-Y%,
a non-zero $\alpha$. Price coefficients $\beta_y$ for different income groups are identified by within-plan, over time variation in subsidy rates and plan premiums. The traditional price-endogeneity issue is less of a concern in CommCare. Individuals with incomes below the federal poverty line accounted for more than 40% of the total CommCare enrollees every year. Health insurance for this group was always fully subsidized in CommCare. Therefore, the effective price sensitivity for below-poverty consumers is zero, and the underlying $\beta_y$ for this group is not a coefficient that I estimate. But due to its large size and due to the particular institutional details of regulations regarding this income group on CommCare, I believe there is enough evidence that most (if not all) of the variation in insurer premiums in my data stems from insurers’ strategies regarding this group. Therefore, there should be little concern about premium variation being endogenous to demand from the other income groups, for which I estimate premium coefficients $\beta_y$. For more detail on this, see the online appendix.

**Results and Interpretation:** Table 3 summarizes the results of the multi-nomial logit estimation of the plan demand model. An average CommCare enrollee with an income between 100% and 200% of the federal poverty line is estimated to be almost twice as sensitive to the after-subsidy premium paid as an average enrollee with an income between 200% and 300% of the federal poverty line. Brand effect estimates indicate that except for Celticare, the brand values for all the CommCare plans are close to one another. An average CommCare enrollee with an income between 100% and 300% of poverty would be willing to pay about $30/month less (in after-subsidy premium) for Celticare than she would for BMC, if the two plans offered the same network of hospitals. Similarly, the average above-poverty CommCare enrollee would value the NHP brand and Network Health brand, respectively, at $8/month and $3/month below that of BMC. This is exactly where the model captures the idea that Celticare is a smaller insurer than the others on CommCare. A smaller brand fixed effect $\Delta_j$ implies that a plan would get a smaller market share even if it offered the same hospital network as the other insurers. This, as will be discussed later, hinders Celticare from adding hospitals (especially smaller ones) to its network since the number of new enrollees that those hospitals would bring Celticare is too few to justify the corresponding fixed cost.

Finally, the network utility coefficient $\alpha$ is estimated at 0.76. This magnitude for $\alpha$ implies that, an average CommCare enrollee with an income between 100% and 300% of poverty would be willing to pay almost 21$ more
in after-subsidy premium for Celticare, if Celticare improved its network of hospitals to one similar to NHPs.

### Plan Demand Estimates

<table>
<thead>
<tr>
<th></th>
<th>coeff.</th>
<th>std. error</th>
</tr>
</thead>
<tbody>
<tr>
<td>network expected utility</td>
<td>0.76***</td>
<td>(0.14)</td>
</tr>
<tr>
<td>price coefficients</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-100 % poverty (omitted, no premiums)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>100-200 % poverty</td>
<td>-0.0413***</td>
<td>(0.0028)</td>
</tr>
<tr>
<td>200-300 % poverty</td>
<td>-0.0198***</td>
<td>(0.0021)</td>
</tr>
<tr>
<td>brand fixed effects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMC (omitted category)</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td>Celticare</td>
<td>-1.137***</td>
<td>(0.041)</td>
</tr>
<tr>
<td>Fallon</td>
<td>0.012***</td>
<td>(0.004)</td>
</tr>
<tr>
<td>NHP</td>
<td>-0.296***</td>
<td>(0.002)</td>
</tr>
<tr>
<td>Network Health</td>
<td>-0.117**</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Num. enrollments</td>
<td>381901</td>
<td>—</td>
</tr>
<tr>
<td>pseudo $R^2$</td>
<td>0.704</td>
<td>—</td>
</tr>
</tbody>
</table>

std errors in parentheses, *: $p < 0.1$, **: $p < 0.05$, ***: $p < 0.01$

Table 3: Plan Demand Model

### 4.4 Step 3: Estimating Insurers’ Non-Inpatient Costs

Variable costs to insurers have two main components. The first component is inpatient costs, which are the reimbursements $t_{ij}$ that the insurer pays to its in-network hospitals for the care that they provide to its enrollees.

---

A last note on my plan demand model is that it does not account for potential inertia in plan choice by consumers. Even though it would be reasonable to believe that such inertia does exist, I think abstracting away from the complexities that it would add to my model does not drastically change the results. This is because in the plan demand estimation, I use data only from fiscal years of 2012 and 2013. And when I use these estimates (which do not account for inertia) to project market shares for the fiscal year of 2011, my predictions are close to the observed shares. Given that substantial changes happened to some plans between 2011 and 2012 as well as between 2012 and 2013, I believe that the model doing well on predicting 2011 market shares suggests that abstracting from inertia is not a far from reasonable.

---

33
The second component is the non-inpatient cost $\xi_j$ in (10). In the online appendix, I detail how I measure reimbursement rates $t_{ij}^*$ using the medical claims file of the MA-APCD. In the main text, I treat them as data. In this section, I back out the non-inpatient costs $\xi_j$.

**Estimation Procedure:** The parameters to back out in this step are insurers’ non-inpatient costs $\hat{\xi}_j$ for FY2011. This is done by imposing the Nash Bertrand equilibrium assumption on the premium-setting. That is, no insurer $d_j$ should be able to do better by changing its premium:

$$\pi_{d_j}(G^*, T^*, P^*) \geq \pi_{d_j}(G^*, T^*, (p_j, p_{-j}^*))$$  \hspace{1cm} (19)

Insurer $d_j$’s non-hospital cost per member per month is estimated by finding the $\xi_j$ to make this optimality condition hold.

**Identification:** Unfortunately, the identification is only partial in this step of the estimation process. In FY2011, the before-subsidy premiums bid by all CommCare insurers were enforced by regulation to be no more than $425$/month and no less than $404$/month. Thus, all that the identifying assumption (i.e., equation (19)) implies is that for every insurer $d_j$, the observed premium $p_j^*$ in the data would do weakly better than any other $p_j \in [404, 425]$. Given that all insurers’ premiums were exactly at either of the two extremes of this continuum, equation (19) only gives us bounds on $\xi_j$ values rather than point estimates.

**Results and Interpretation:** Table 4 presents the bounds backed out on non-inpatient costs $\xi_j$ using (19).

**Calibration:** Given that I only have bounds on $\xi_j$, I need to calibrate the non-inpatient costs. I calibrate the vector $\hat{\xi}$ exactly at the boundaries of the inequalities. That is, $\hat{\xi} = (284, 275, 281, 276)$. I believe this is a reasonable calibration for multiple reasons. First, the fact that Celticare covers, on average, more expensive hospitals than the other three plans (see table 1) can also push Celticare’s outpatient cost further upward compared to others. Therefore, given that, by table 4, the upper bound on $\xi_{\text{Celticare}}$ is below the lower bound on $\xi_j$ for the other plans, calibration at the boundary seems not

<table>
<thead>
<tr>
<th>Bound on $\xi_j$ ($$/month/person)</th>
<th>BMC</th>
<th>Celticare</th>
<th>NHP</th>
<th>Network Health</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of Bound</td>
<td>Lower</td>
<td>Upper</td>
<td>Lower</td>
<td>Lower</td>
</tr>
<tr>
<td>Bound on $\xi_j$ ($$/month/person)</td>
<td>284</td>
<td>275</td>
<td>281</td>
<td>276</td>
</tr>
</tbody>
</table>

Table 4: Upper or lower bounds on non-inpatient costs $\xi_j$
far from reasonable. Also, even though non-inpatient costs $\xi_j$ may have varied over time, looking at the pricing decisions of insurers in fiscal years other than 2011 can still be suggestive regarding how reasonable the decision to calibrate $\hat{\xi}$ at the boundary is. In fiscal year 2010, the premium floor and cap were $381/month and $391/month respectively. The fact that BMC, NHP, and Network Health could all profitably charge no more than $391/month (i.e., $34/month below their prices in FY2011) suggests that calibrating their $\xi_j$ at the lower bounds is a good working assumption. In addition, during fiscal years of 2012 and 2013 when CommCare had set the premium floor at the much lower levels of $360/month and $354/month respectively, and put in place a procurement process to encourage insurers to price lower, Celticare was not always among the two cheapest plan on CommCare.\footnote{For more detail on CommCare procurement rules for fiscal years of 2012 and 2013 and the insurers’ premiums during those years, see the online appendix.} This suggests that Celticare must not have had a $\xi_j$ much lower than the other plans, which is consistent with my calibration.

4.5 Step 4: Estimating Insurer Fixed Costs and Bargaining Parameters

Modifying the Model: Before I perform the estimation in this step, I make modifications to the model to avoid computational burden for the estimation of fixed costs and bargaining parameters and, more importantly, for the simulation of counterfactual scenarios. First, instead of the whole state of Massachusetts, I concentrate on the Greater Boston Area which I define by zip codes that are at most 30 miles far away from 02114, a zip code in downtown Boston. There are 28 hospitals in this area. Also, Fallon was not active in CommCare in this area. So, I am left with BMC, Celticare, NHP, and Network Health. Second, I coarsen the set of consumer bins $\kappa$ by aggregating the zip codes into 38 different locations. I also aggregate all DRG weights into a single one\footnote{This aggregation abstracts away from heterogeneity in the severity of different admission events. I decided for this aggregation since trying it on the early versions of the model did little to change the results.}. Third, I simplify the way hospital-insurer pairs $u_i$ and $d_j$ anticipate how premiums respond to the outcome of their bargaining over reimbursement rate $t_{ij}$. I assume that when negotiating, the pair take as given the premiums set by all insurers other than $d_j$. That is, the pair assume that only $d_j$ will adjust its premium optimally after the
outcome of the bargaining between the two is determined. This assumption substantially simplifies the computation as it will not involve computing a complete Nash Bertrand equilibrium for insurer premium-setting for every candidate negotiated $t_{ij}$. In spite of the massive computational advantage, this assumption does capture the first-order effect of premium response to bargaining outcomes. It only leaves out how other insurers respond to the premium change by $d_j$.\footnote{This assumption still retains the feature that at equilibrium market outcome $(G^*, T^*, P^*)$, premiums $P^*$ are a Nash Bertrand equilibrium among insurers, given $G^*$ and $T^*$.} This type of simplifying assumption has been made in the literature to ease the computation of models of bargaining on two sided markets. In fact, Ho and Lee (2017\textit{b}) assume that when bargaining, hospital $u_i$ and insurer $d_j$ take as given all of the premiums, including $p_j$. The assumption in Ho and Lee (2017\textit{b}) simplifies the computation much further than the assumption I impose. Nevertheless, as I will discuss in the online appendix, it would not be suitable to a model of bargaining with endogenous network formation such as the model in this paper, as it would lead to wide ranges of multiple equilibria in many cases.

The fourth modification to the model not only simplifies the computation, but is also rooted in the institutional details of the market. I assume hospitals do not bargain individually, but as systems. The participation of hospitals owned by the same hospital system across plans’ hospital networks is highly correlated with one another with full correlation in most cases. Therefore, I assume that it is the whole system, denoted $s$, that negotiates with each insurer $d_j$ over rates and that the outcome of the negotiation is the participation of either all of the system’s hospitals $u_i$ in $d_j$’s network or none of them. This partitions the 28 hospitals in the Greater Boston Area into 16 systems.\footnote{If, within an actual system (i.e., joint ownership of hospitals), the network participation of some hospitals do not fully correlate with one another, I consider them separate systems. For instance, Faulkner Hospital is owned by the Partners system, which means that it is in the same system as MGH. Nevertheless, I consider them to be in separate systems, as they do not always appear in the same hospital networks.} I also assume that in the case of full participation, the prices negotiated between insurer $d_j$ and all hospitals $u_i$ in system $s$ are fully correlated. That is, I construct a fixed “base-rate matrix” $T_{\text{base}} = \begin{bmatrix} t_{ij}^{\text{base}} \end{bmatrix}_{m \times n}$ and assume that the negotiated prices between insurer $d_j$ and all hospitals $u_i$ in system $s$ can only take the form of $t_{ij} = t_{ij}^{\text{base}} + z_s$.\footnote{For every $ij$ link that does appear in the network structure based on my FY2011 data,}
fact over $z_s$. This assumption brings the dimensionality of each bargaining process from the size of system $s$ to 1, without substantially affecting the outcomes. After implementing all these four modifications, I turn to the estimation of fixed costs and bargaining parameters.

**Estimation Procedure:** The procedure closely follows [Gowrisankaran, Nevo and Town (2013)](GNT) (henceforth GNT.) They jointly estimate bargaining parameters and hospital marginal costs. I go through a similar process for bargaining parameters and fixed costs $f_{sj}$. In a similar spirit to GNT, I make assumptions on the costs and bargaining parameters.

For the bargaining parameters, I divide all of the hospital systems into two categories: “star” and “non-star.” I consider two hospital systems to be stars. First, Partners, which includes Massachusetts General Hospital and Brigham and Women’s Hospital. The second one is Tufts Medical Center. These systems are the highest ranked in Massachusetts according to U.S. News and World Report’s 2016-2017 rankings of best hospitals in Massachusetts. Another reason for considering these two systems to be star systems was that they seemed to have charged abnormally high to all insurers according to my measurement of $T^*$. I then assume that bargaining parameter $\gamma_{sj}$ equals $\gamma_{\text{star}}$ for all $j$ when $s$ is one of the two star systems, and it equals $\gamma_{\text{nonStar}}$ otherwise.

For the costs, I specify $f_{sj}$ as linearly determined by observable characteristics of hospital systems and insurers. I choose total hospital-system bed sizes as well as insurer fixed effects for those characteristics:

$$f_{sj} = \Omega \times \chi_s + F E_j + v_{sj}$$  \hspace{1cm} (20)

I set $t_{ij}^{\text{base}}$ to the value of the estimated reimbursement rate $t_{ij}^*$ corresponding to that link, which, as mentioned before, I measure in the online appendix. For all other links $ij$, I set $t_{ij}^{\text{base}}$ to hospital $i$’s marginal cost $c_i$.

To see why this assumption does not have large effects on the outcome, observe that increasing some $t_{ij}$ and decreasing some $t_{ij}'$ would have little net effect on either the insurer profit or the hospital system profit (which is the sum of the profits to the individual hospitals forming that system). Therefore, the bargaining here could be thought of as the insurer’s attempt to get discounts from all of the individual hospitals in the system, and the system’s attempt to charge the insurer more for the service of all of the hospitals.

In that ranking, Massachusetts General Hospital, Brigham and Women’s Hospital, and Tufts Medical Center are ranked first, second, and fourth respectively. Beth Israel Deaconess Medical Center was ranked third, above Tufts Medical Center. But I decided to not consider its parent system (i.e., Care Group) a star system since it also included many much lower ranked hospitals such as Beth Israel Deaconess Needham and Beth Israel Deaconess Milton, and especially Mount Auburn Hospital.
The above gives the following econometric error term:

\[ \nu_{sj} = f_{sj} - \Omega \times \chi_s - FE_j \]  \hspace{1cm} (21)

The next GNT-counterpart step is to use the moment condition

\[ E[\nu_{sj}|G^*, T^*, P^*, C, \xi, D, \sigma] = 0. \]

More specifically, I find the values of \((f_{sj}, \gamma_{sj})\) that minimize the magnitude of the econometric error, subject to (i) the equilibrium conditions of the model, and (ii) the assumptions imposed in this subsection (i.e., equation (20) and \(\gamma_{sj}\) taking only one of the values \(\gamma_{star}\) and \(\gamma_{nonStar}\)). The solution to this minimization problem will yield the estimates \(\hat{f}_{sj}\) of fixed costs and \(\hat{\gamma}_{sj}\) of bargaining parameters.

Aside from estimating fixed costs instead of marginal costs, there is another difference between the estimation process in this paper and GNT. In GNT, the imposed set of equilibrium conditions is NiN bargaining. In this paper, there are two deviations from that. First, network-formation conditions are also imposed. That is, for any link \(sj\) that is absent from the equilibrium, a lower bound is imposed on the corresponding \(f_{sj}\). Second, the bargaining formulation in this paper is different from NiN. These two differences make the process of solving for the minimizer \((\hat{f}_{sj}, \hat{\gamma}_{sj})\) substantially more complex. Nevertheless, the nature of the minimization problem is the same as in GNT. The outcome of this procedure will be the tuple \((\hat{f}_{sj}, \hat{\Omega}, \hat{FE_j}, \hat{\gamma}_{star}, \hat{\gamma}_{nonStar})\). One dimension along which solving the optimization problem is harder here is that, in GNT, the NiN bargaining conditions fully pinpoint the costs conditional on bargaining parameters. Therefore, the optimization problem is effectively over bargaining parameters. But here, for any pair of values \((\gamma_{star}, \gamma_{nonStar})\), we get a set of values for the fixed costs \(f_{sj}\) (e.g., for absent links we only get lower bounds.) Thus, the process has two steps. First, for each candidate pair \((\gamma_{star}, \gamma_{nonStar})\), I find the fixed costs that minimize the econometric error. Then I minimize over bargaining parameters to obtain \((\hat{\gamma}_{star}, \hat{\gamma}_{nonStar})\). The first stage of this process is discussed in detail in the appendix. The process that I develop for the first stage of this two-stage optimization procedure is called the “Regression

A question that may arise here is: The bargaining formula for any \(sj\) is the minimum between the Nash Bargaining rate and the highest safe one. Which one should be imposed? The answer is: That is determined as part of the optimization problem. Equation (5) implies (i) \(t_{ij}^* \leq t_{ij}^{NB} (G, T, \gamma_{ij})\), (ii) \(t_{ij}^* \leq \hat{t}_{ui} (G, T, d_j)\), and (iii) that either (i) or (ii) holds with equality. Each of these is a constraint to the optimization problem. More details are given in the appendix.
Fixed-Point” algorithm, which I believe could expand the use of the GNT estimation procedure. In the online appendix, I also give more details on why I chose this GNT-based approach over other alternatives such as MLE or moment inequalities.

**Identification:** Identification of all of the parameters come jointly from the structural assumptions of the model described in section 2, as well as from the moment condition in (20). Fixed effects \( FE_j \) are roughly identified by average differences among different insurers \( d_j \) in the bounds on their respective \( f_{sj} \) by the structural assumptions. For instance, the imposition of no gains from trade by definition 1 on links \( s_j \) with \( g^*_{sj} = 0 \) is expected to imply a lower bound \( \underline{f}_{sj} \) on \( f_{sj} \). If such implied lower bounds for insurer \( d_j \) tend to be larger than those implied for insurer \( d_{j'} \), then \( FE_j \) can be larger than \( FE_{j'} \). Also given the structure of the model, the bargaining formulation (8) implies an upper bound \( \bar{f}_{sj} \) on \( f_{sj} \). A similar argument applies when \( \bar{f}_{sj} \) tends to be larger for \( d_j \) than for \( d_{j'} \). The intuition for the identification of \( \hat{\Omega} \) is also similar, though the variation this time is within insurer but across hospital systems.

The bargaining parameter for star hospital systems \( \gamma_{\text{star}} \) is identified by searching for the \( \gamma_{\text{star}} \) value that makes the predicted fixed costs \( f_{sj} \) for the Partners and Tufts systems as close as possible to what a linear model of fixed costs with \( \hat{\Omega}, FE_j \) would predict. For instance, if the fixed effects \( FE_j \) and size effect \( \hat{\Omega} \) are all positive, but with \( \gamma_{\text{star}} = \frac{1}{2} \) the implied fixed costs for all or most \( s_j \) with \( s \in \{\text{Partners, Tufts}\} \) and \( g_{sj} = 1 \) are negative, then we expect that the “true” \( \gamma_{\text{star}} \) must have been larger than \( \frac{1}{2} \). That is, the insurers must have accepted higher reimbursements from those star hospital systems, not because of very low fixed costs of inclusion, but because those systems have higher bargaining parameters. Bargaining parameters \( \gamma_{\text{nonStar}} \) are also identified based on a similar logic.

A particular feature of my model that further helps to separately identify fixed-costs of inclusion from bargaining parameters is endogenous network formation. Some of the equilibrium conditions of the model only restrict fixed costs \( f_{sj} \) and are invariant to bargaining parameters \( \gamma_{sj} \). For instance, lower bounds \( \underline{f}_{sj} \) for those \( s_j \) with \( g^*_{sj} = 0 \) are completely independent of the whole bargaining-parameters matrix. So, very small values of \( \gamma_{\text{star}} \) and \( \gamma_{\text{nonStar}} \) would imply very low and sometimes negative upper-bounds.

\(^{45}\) Also, given that networks of different plans can be different, sometimes \( \bar{f}_{sj} \) can be compared to \( \underline{f}_{sj} \). Similar arguments apply here.
that are far below \( f_{sj} \), will not do well in minimizing the error in regression equation (20), and hence are not chosen by the estimation algorithm.

**Results and Interpretation:** Table 5 summarizes the results. The two hospital systems I consider star are estimated to have full bargaining power \( \hat{\gamma}_{\text{star}} = 1 \) when negotiating with insurers.\(^{46}\) I estimate that non-star systems have a bargaining power of \( \hat{\gamma}_{\text{nonStar}} = 0.72 \). The fixed costs of inclusion \( f_{sj} \) are estimated to be increasing with the total bed size of the hospital system covered. The slope of this increasing relationship is estimated at \( \hat{\Omega} = \$223\text{K/year/100 beds} \). This magnitude, along with the fixed effects \( FE_j \) reported in the table, implies that such fixed costs sum up to as large as about a third of insurer profits (or, put differently, a fourth of insurer profits before accounting for the fixed costs themselves). Observe that the fixed costs estimated for Celticare are not higher than those for the other insurers. So, the rationization of Celticare’s narrower network than the other plan is not taking place by simply assigning higher fixed costs \( f_{sj} \) to Celticare. As mentioned before, an important role here is played by Celticare’s low brand fixed effect \( \Delta_{j} \) in the estimated plan demand model.\(^{48}\)

A final observation is that the magnitude of the estimated fixed costs seem rather high. One reason for this can be that the Nash Bertrand assumption on the premiums tends to imply overly high insurer markups, thereby underestimating the non-inpatient costs. Such underestimation would overstate

\(^{46}\)The analysis of these lower bounds from the network formation structure was also the reason why I chose how to “split” the effects between the bargaining model and the fixed costs model. In the early versions of the model, I also tried specifications where the fixed costs model included a star dummy but no size effect and the bargaining model had a size effect. That approach led, on average, to a larger diversion in \( f_{sj} - FE_j \) for a fixed \( j \) across \( s \) than my chosen specification does, which means it does worse on matching the moments.

\(^{47}\)An estimated bargaining parameter of \( \hat{\gamma}_{\text{star}} = 1 \) might seem too high since it seems to give all the surplus from the contract to the hospital. Two notes about this issue are worth making. First, all the surplus does not necessarily go to the hospital since the insurer is also allowed to make threats of replacement. Second, and more importantly, standard plan demand models in the literature (which I also use in this paper) do not capture consumer risk aversion as a source of bargaining leverage for star hospitals. Therefore, this leverage in my estimation process is picked up by an extremely high estimated bargaining parameter \( \hat{\gamma}_{\text{Star}} \). See the online appendix for more detail on this.

\(^{48}\)In fact, I ran a counterfactual simulation using the estimated parameters of the model, except that I set all of the brand effects \( \Delta_{j} \) to zero to assume away brand heterogeneity. In the simulated equilibrium, Celticare and Network Health ended up having 10 hospital systems (out of the total 16) covered in their networks.
the value of new enrollees and, hence, the value of including an additional hospital system, to insurers. This makes higher fixed costs necessary to rationalize insurers’ choice of network.

### Fixed Costs and Bargaining Parameters

<table>
<thead>
<tr>
<th></th>
<th>coeff.</th>
<th>std. error</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>insurer fixed effects (K$/year)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMC</td>
<td>1146***</td>
<td>(180)</td>
</tr>
<tr>
<td>Celticare</td>
<td>788***</td>
<td>(177)</td>
</tr>
<tr>
<td>NHP</td>
<td>415**</td>
<td>(181)</td>
</tr>
<tr>
<td>Network Health</td>
<td>636***</td>
<td>(177)</td>
</tr>
<tr>
<td><strong>effect of size (K$/year/100 beds)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ω</td>
<td>223***</td>
<td>(91)</td>
</tr>
<tr>
<td><strong>bargaining parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>γ_{star}</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>γ_{nonStar}</td>
<td>0.72</td>
<td></td>
</tr>
</tbody>
</table>

std errors in parentheses, *: p < 0.1, **: p < 0.05, ***: p < 0.01

Table 5: Estimation of fixed costs of inclusion and bargaining parameters

### 5 Counterfactual Analysis

I now simulate the effects of a range of network adequacy regulations using my estimated model on CommCare for the Greater Boston Area and FY2011. I examine network adequacy regulations in the form of mandating all insurers in the market to cover at least X% of the hospital systems in the Greater

---

49The issue of Nash Bertrand conditions overstating markups has been mentioned in the literature before. [Ho and Lee (2017)] address it by assuming that insurers perceive demand to be more elastic than it actually is. The estimates in this paper can be thought of as another alternative: Insurers failing to perceive the full relevance of their non-inpatient costs when setting premiums but do take them into account when forming networks, or equivalently, insurers implicitly taking part of their variable costs to be fixed. I, hence, do not expect the magnitude of such fixed costs to drastically affect the results that I get in my counterfactual analysis as long as they help to capture the heterogeneity in profitability levels of different hospitals to the insurer that are not otherwise captured, since these differential profitability levels are key to both which hospitals are included in the network and how strong they can be against each other as threats of replacement.
Boston Area, for a range of values of $X$. To model this mandate, I just include an additional term in the insurer-profit-function in (10), representing a fine that the insurer has to pay if it falls below the $X$% requirement:

$$
\pi^{NA}_{d_j}(G, T, P) = \pi_{d_j}(G, T, P) - \eta \Gamma_j(G, X)
$$

where $\pi_{d_j}(G, T, P)$ is insurer $d_j$’s profit function from equation (10), and $\Gamma_j(G, X)$ is the number of hospital systems by which insurer $d_j$ is short of fulfilling the $X$% requirement (e.g., $\Gamma_{\text{Celticare}}(G^*, 0.5) = 16 \times 0.5 - 4$ because Celticare covers 4 hospital systems and is, hence, 4 hospital systems short of 8 which would be 50% of the total 16 systems). Also $\eta$ is the amount of fine an insurer has to pay for each of those missing hospital systems. The exact amount of the fine $\eta$ does not play a role in determining the result of the counterfactual simulation. It only has to be large enough to ensure that no insurer would prefer going below $X$% to staying weakly above it. I then simulate the new equilibrium of the model again, using the parameters estimated in section 4 and using (22) instead of (10) for insurer profit.

The rest of this section is organized as follows. Section (5.1) analyzes the simulated response of CommCare to network adequacy regulations and

---

50 Other types of network adequacy regulations to consider could be based on hospital counts or bed counts rather than hospital-system counts. Examining those regulations can be done using my model, but it would require first modifying the bargaining formula in (8) to capture threats of replacing multiple in-network hospital systems by multiple out-of-network ones. For instance, if an in-network hospital system $u_s$ has five hospitals and no out-of-network system is as large, the insurer needs to be able to threat $u_s$ to replace it, for instance, by two other systems $u_{s'}$ and $u_{s''}$ consisting of two and three hospitals respectively. I believe the main difference between the hospital-system based regulation in my model and, say, a total-beds-based one would be that a system based regulation will give more bargaining leverage to hospital systems with smaller total bed sizes (and, thereby, less leverage to systems with more beds) compared to a bed-based regulation, since a small system counts just as much as a larger system in terms of helping the insurer abide by the minimum mandated network size.

51 A more natural way of modeling the fine would be having a constant fine for being below 50% no matter how much below. With that formulation, however, my model would not capture Celticare’s incentive to expand its network of hospitals. The reason is that the deviations allowed in my model are dropping, adding, or replacing one hospital system. No single such deviation can take Celticare from covering 4 systems to 8 systems. Therefore, if the fines for any degree of non-compliance with the requirement is the same, my model will not predict a change in Celticare’s behavior.

52 If the bargaining model was NiN instead of bargaining with threats of replacement, then $\eta$ would play a significant role, with $\eta \rightarrow \infty$ implying $t^{*}_{ij} \rightarrow \infty$, which I do not find reasonable.
summarize the generalizable lessons learned about policy making. It also discusses what results NiN would imply. Section (5.2), performs consumer-welfare analysis. In section (5.3), I discuss how differently we should expect the ACA exchanges to respond to the regulation than CommCare does.

5.1 CommCare’s Response to Network Adequacy Regulations

There are at least three important policy-making lessons learned from simulating the regulations on CommCare. First, reimbursement rates and premiums can drastically rise under stringent network adequacy regulations. Second, if the narrow network formed by an insurer consists mainly of high-cost hospitals, then a network adequacy regulation may indeed reduce the average reimbursement rate paid by that plan to its in-network hospitals. Third, even though network adequacy regulations are designed to expand narrower hospital networks, they sometimes end up shrinking the broader ones. Figures below summarize these lessons. In these figures, the horizontal axis shows the minimum percentage X mandated to cover, and the vertical one corresponds to network size (in percentage of systems covered,) average reimbursement rate, or premium, depending on the figure. Below, I discuss these lessons in more detail.

As shown in figure (6), whenever the is binding for Celticare, Celticare responds by including as many hospital systems as required.

Lesson 1: As shown in figures (7) and (8) reimbursement rates and premiums rapidly increase towards the end of the spectrum of network adequacy regulations. This has two reasons. First, as mentioned earlier, with more stringent regulations, the insurer’s bargaining position is severely compromised due to sorting of hospitals. The insurer cannot threaten to drop any hospital system; and on top of that, it can only threaten to replace with systems that were so high-cost or low-quality that they did not make it into the network even under the heavy regulation. Of course such systems cannot provide Celticare with strong enough threats of replacement against its in-network hospital-systems, which leads the prices to rise.

The second reason behind the price acceleration is a nuanced self-reinforcing loop: Under the regulation, each hospital system charges Celticare for the differential profitability from the best alternative, currently-out-of-network, hospital system. A premium-increase affects hospitals more than it impacts
Figure 6: Effects of network adequacy regulations on Celticare’s network size

Figure 7: Effects of network adequacy regulations on Celticare’s payments to in-network hospitals
Figure 8: Effects of network adequacy regulations on Celticare’s premium

Figure 9: Effects of network adequacy regulations on Celticare’s premium
Celticare. It reduces Celticare’s enrollee volume but this reduction is (at least partially) offset by the margin increase that directly results from the premium rise. Celticare’s hospitals, however, only undergo the volume decrease. Therefore, the differential profitability levels of hospital systems to Celticare do not change much; but given the shrunken volumes, a hospital system needs a higher reimbursement rate to charge for the difference. As a result, reimbursement rates rise, to which Celticare best-responds by further raising its premium. This feeds back into raising reimbursement rates even more, and so on. The bottom line is, the magnitude of price consequences may not be smooth as a function of the extent of the regulation.

**Lesson 2:** As depicted in figures (7) and (8), for mid-range regulations, the average reimbursement-rate per admission paid by Celticare to its in-network hospitals in fact decreases. This is because, absent regulation, Celticare covers high-cost and prestigious hospitals such as MGH and BWH and excludes cheaper ones. Thus, once forced to add hospitals, it includes some that have much lower marginal costs of providing care than its originally in-network hospitals did. These new hospitals tend to charge less than hospitals like MGH or BWH, lowering the average reimbursement rate. To see that this is indeed what explains the rate-decrease for mid-range regulations, observe the orange dashed curve which represents the average marginal cost of providing inpatient care calculated for Celticare’s in-network hospitals. It decreases as the regulation intensifies. Also observe the blue dotted curve which shows the average rate charged by the four systems included in Celticare’s original (pre-regulation) network. This average rate, unlike the overall average rate, does not decrease as the regulation becomes more stringent.

**Lesson 3:** Figure (9) depicts the response of NHP’s network size to the regulation. As depicted there, NHP responds even before the regulation becomes binding for it (i.e., before the graph hits the 45-degree line, dashed-depicted in the figure.) In other words, NHP is responding to the response of Celticare (and then BMC) to the regulation by dropping some hospital systems from its network. This happens because those hospitals, who get better deals from narrower-network plans as a result of the regulation, now bargain more aggressively with NHP. In some cases it pays off for NHP to raise their rates and keeps them in-network; and in other cases they drop out to steer their more loyal patients to the other plans that now are paying higher. The bottom-line is: aside from the price consequences, even the network-breadth consequences of network adequacy regulations is may be adverse.
Table 6: Illustration of Celticare’s average reimbursement rate per admission under different penalty values, comparing NiN with and without threats for replacement.

Before turning to welfare analysis, I discuss the issues one would face if one analyzed this regulation using a NiN framework. Table (6) illustrates, for comparison, the behavior of a NiN model under the regulation. Consider a regulation with \( X = 37.5 \) (i.e., a minimum of 6 systems.) The table shows average reimbursement rate under the equilibrium network paid by Celticare to its in-network hospitals under that regulation with different penalty values \( \eta \). As mentioned before, in my proposed model, once \( \eta \) is large enough to ensure the minimum network size can be imposed, its exact value is immaterial. However, using NiN, the higher \( \eta \), the more hospital take advantage of the insurer in bargaining and charge rates that are many times as large as a typical rate in this market.

5.2 Welfare Analysis

I measure the welfare effects of network-adequacy regulation in terms of dollars that would need to be subtracted from all premiums in the original market outcome in order to generate the same average consumer utility. To illustrate, if the impact of a particular regulation level is $7, it means that the welfare effect of this regulation is equivalent to reducing the premiums in the original market-outcome –i.e., that observed in the data– from (425,404,425,425) to (418,397,418,418). Figure (10) shows the welfare sim-

\[ \text{Note that table (6) finds the NiN rates taking as given the equilibrium network implied by my modified model with threats of replacement. It is not possible to find what network NiN itself would imply under network adequacy regulations. That is because at the boundary of the regulation, the rates go high, as suggested by the table, leading Celticare to add hospitals to steer the patients to lower charging hospitals (the new hospital cannot charge a lot since it is not pivotal.) But once the new hospital is added, no hospital is pivotal anymore, leading the prices to drop, and, hence, leading Celticare to drop a hospital. This makes all hospitals pivotal again, and so on.} \]
According to figure (10), the optimal regulation level would be 81.25 (i.e., a minimum of 13 hospital systems.) That is the point where the network has expanded but the rapid price increase has not kicked in yet. Figure (11), however, shows that the overall welfare results are expected to be different than the Celticare-only results in at least two ways. First, the positive welfare identified due to their full subsidization.

In fact, in figure (10), welfare always increases in X while X < 81.25 with the exception of the point X = 25. That is the point where the regulation mandates the same minimum size as Celticare’s original, pre-regulation, size. That will not expand Celticare’s network but has price consequences, since previously Celticare could threaten to go below %25 but now it cannot. Thus the overall welfare effect is negative.

Figure 10: Welfare effects of the regulation if Celticare (but no other insurer) responds.

Figure 11: Overall welfare effects of the regulation
effect over mid-range regulations is now a somewhat neutral one with some positive and some negative points. Second, the optimal regulation is the most stringent one. This latter feature is present simply because all plans besides Celticare charge $425/month/person, and, hence, cannot pass on the increased reimbursement rates to consumers. The first features is more generalizable to other insurance markets. It shows that the downsize of broader networks in response to network adequacy regulations could in fact be a first order effect deserving attention in policy design.

5.3 Notes on Policy Design for the ACA Exchanges

This section discusses the takeaways from this analysis for policy design regarding network adequacy regulations in the ACA exchanges. The basic economic forces introduced in section (5.1) can be present in any health insurance market and are, hence, generalizable lessons. The exact welfare results, on the other hand, are more CommCare specific. Next, I outline a non-exhaustive list of the differences between CommCare and the ACA exchanges which can lead to differences in how desirable network adequacy regulations are in these two markets. I also discuss how consumer inattention to networks can affect the optimal regulation policy.

The first difference is in the subsidy structures. In CommCare, subsidies have a large proportional component (e.g., below poverty consumers pay $0/month no matter the premium.) In the ACA, for each consumer, the amount of subsidy is fixed. So, the ACA consumers feel all the price differences across plans, which renders them more price-sensitive. Thus, high-cost narrow-network plans such as Celticare are expected to be less commonplace in the ACA exchanges. This implies the price decrease from mid-range regulations may not be expected to happen in the ACA.

The second difference is that in the ACA there are non-trivial co-pays and co-insurance payments. This, unlike in CommCare, makes the ACA customers directly sensitive to reimbursement rates, possibly mitigating the price consequences of network-adequacy regulations. The third difference is, unlike CommCare, there are no premium caps in the ACA. Therefore, the increased reimbursement rates in response to the regulation can be much more easily passed on to consumers.

Aside from CommCare-ACA differences, there another critical issue for policy design. There is evidence that consumers, at the time of buying insurance plans, pay less attention to the provider-networks; and become dis-
Narrow networks in ACA have lower-cost hospitals
ACA has cost sharing subsidies
ACA does not involve premium caps
Consumers are inattentive to network breadth

Table 7: Policy design notes about the ACA not captured in this paper

<table>
<thead>
<tr>
<th>Note</th>
<th>Implication for regulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Narrow networks in ACA have lower-cost hospitals</td>
<td>against regulation</td>
</tr>
<tr>
<td>ACA has cost sharing subsidies</td>
<td>for regulation</td>
</tr>
<tr>
<td>ACA does not involve premium caps</td>
<td>against regulation</td>
</tr>
<tr>
<td>Consumers are inattentive to network breadth</td>
<td>for regulation</td>
</tr>
</tbody>
</table>

satisfied once they are in need of healthcare and their preferred provider is out-of-network (see, for instance, McKinsey (2014).) Protecting consumers against their own miscalculation has been a major premise for network adequacy regulations. The way this issue can be captured in my framework is by assuming that there is a “true” network-expected-utility coefficient $\alpha_{true}$ which is larger than $\alpha = 0.76$, the coefficient implied by the consumers’ plan choice behavior. The fraction $\frac{\alpha_{true}}{\alpha}$ would be a measure of consumer inattention (at the time of plan choice) to provider networks. This fraction cannot be directly estimated, but surveys might help find a reasonable range for it. My model can then be used to find the minimum inattention level that is necessary to make a certain level of regulation worthwhile for consumers\footnote{This can be done by replacing $\alpha$ with $\alpha_{true}$ in the welfare simulations.} and check if this minimum is within the identified range.

Table below summarizes the key policy-analysis points discussed in this subsection.

6 Conclusion and Future Research

In this paper, I developed a model of insurer-provider markets with three important features. First, it endogenously captures the formation of hospital networks, bargaining between hospitals and insurers over reimbursement rates, and premium setting by insurers. Second, in formulating the bargaining process, my model improves upon a standard model called Nash-in-Nash by allowing for the possibility that when bargaining with hospitals over rates, insurers not only threaten to drop hospitals from the network, but also to replace them with currently out-of-network hospitals. This helps capture the Bertrand-type competition among hospitals for network inclusion and the fact that insurers can play hospitals off against each other and get lower
prices. The third feature is computational tractability. My model can be estimated and used for counterfactual simulations on relatively large markets. In this paper, I also develop an estimation procedure for the model and apply it to the CommCare market in the fiscal year of 2011 in the Greater Boston Area with 16 hospital systems and four insurers.

I use my model to study by how much network adequacy regulations in CommCare can undermine insurers’ bargaining position against hospitals and, hence, lead to increased reimbursement rates as well as increased premiums. I find that milder regulations lead to only moderate increases and sometimes even reductions in reimbursement rates and premiums on CommCare. However, enforcing insurers to have almost-full networks can lead to drastic price hikes since it depletes the out-of-network hospital pools that insurers use as threats of replacement against in-network hospitals to keep rates down. I point out some differences between the ACA exchanges and CommCare that might cause ACA exchanges respond differently to such regulations than CommCare does.

This framework developed in this paper can be applied beyond network adequacy regulations. An important example is merger analysis. One of the anti-competitive effects of a merger between two hospitals is that an insurer can no longer use them as threats of replacement against each other. Unlike my model, NiN abstracts away from this effect and, hence, under-predicts the magnitude of the anti-competitive effects. More generally, this framework can be applied beyond the insurer-provider market. It can be empirically applied to other two-sided markets (such as TV channels and cable companies, or manufacturers and retailers) to answer questions centered around how network formation and bargaining take place and affect each other.

References


Commission, Health Policy. 2013. “Public Hearings Health Care Cost Trends in the Commonwealth: CeltiCare Health Plan of Massachusetts, Inc. (NAIC 13632).”


Ho, Kate, and Robin S. Lee. 2017b. “Insurer Competition in Health Care Markets.”


McKinsey. 2015. “The emerging story on new entrants to the individual health insurance exchanges.”


Prager, Elena. 2015. “Tiered Hospital Networks, Health Care Demand, and Prices.”


Schmitt, Matt. 2015. “Do Hospital Mergers Reduce Costs?” *Available at SSRN 2599000*.


Shepard, Mark. 2015. “Hospital Network Competition and Adverse Selection: Evidence from the Massachusetts Health Insurance Exchange.”

