

Generalized Insurer Bargaining

Guy Arie* Paul L.E. Grieco[†] Shiran Rachmilevitch^{‡§}

October 29, 2015

Abstract

We incorporate repeated interaction and limits on the number of simultaneous negotiations by the same insurer into the standard multi-lateral insurer-hospital Nash-In-Nash (NiN) bargaining model. This approach is motivated by our finding that under common assumptions, the NiN model predicts a market breakdown with sufficiently high hospital bargaining power. In our proposed model all hospitals that increase surplus join the insurer network. Our generalized model can be estimated as in Gowrisankaran et al. (2015) with one additional parameter – the players’ discount factor. If players are completely impatient, the estimation outcome is the same in both models. We identify the differences in estimation results between the two models and show that mergers that would be approved using the NiN model may be rejected using the general model.

Keywords: Bargaining; Health economics; Insensitivity to prices.

1 Introduction

Multi-lateral bargaining between downstream and upstream providers is central to important policy decisions. Welfare analysis of mergers between TV content dis-

*University of Rochester, Simon Business School: guy.arie@simon.rochester.edu

†Pennsylvania State University: paul.grieco@psu.edu

‡University of Haifa: shiranrach@econ.haifa.ac.il

§This paper extends and replaces the working paper “Intermediary Bargaining for Price-Insensitive Consumers” by Arie and Rachmilevitch. We thank participants in the UECE Lisbon Meetings (2014), IIOC (2015) and Paul Ellickson, Henry Mak, Jeanine Miklos-Thal, Ron Schmidt, Greg Shaffer and Gerry Wedig for helpful comments. We are especially grateful for David Klein for his input on the institutional aspects. All errors are our own.

tributors (e.g. Time Warner Cable merging with either Comcast or Charter), the recent wave of hospital mergers and the rising consolidation of health insurers all rely on the underlying bargaining model.¹ For concreteness, the sequel will focus on the hospital-insurer bargaining setting.

The theoretical foundation for empirical analysis of bargaining settings is the multilateral bargaining model developed in Horn and Wolinsky (1988), which is a Nash equilibrium of two inter-related asymmetric Nash bargaining games as in Harsanyi and Selten (1972) and Kalai (1977). In the sequel, we follow the term coined in Collard-Wexler et al. (2014) and call this the Nash-in-Nash (NiN) model.

However, despite the growing importance of multi-lateral bargaining for the empirical applications mentioned above, the analysis in Horn and Wolinsky (1988) is motivated by a different setting and does not address some of the key features of the healthcare market. Accounting for these features, we find that NiN in the hospital-insurer setting may inflate prices.

Our first result is that, under common assumptions, the hospital equilibrium prices predicted by NiN may be higher than the patients' value for the services for any observed network. Thus, the market either breaks down and all insurers quit or the insurers remain and deliver negative surplus. This is of course undesirable and suggests that a generalization of the estimation model that accounts for the specific market features is important.

Motivated by this result, we develop an alternative Repeated Sequential Nash (hereafter RSN) model. Estimation based on the RSN model can be viewed as a generalization of NiN with an additional discount factor to be estimated. The NiN estimation result is obtained at the limit as the discount factor approaches zero. However, the equilibrium per-service price paid by the insurer is guaranteed to be lower than the value of the service to the insured.

In the NiN bargaining model, each hospital bargains with the insurer holding the contracts with all other hospitals fixed. Price is determined by comparing the network's surplus with and without the hospital. Consider bargaining with some hospital A given the negotiated prices with hospital $B \neq A$. The network's surplus without A depends not only on B 's price but also on A 's patients that would now

¹Crawford and Yurukoglu (2012) and Lee et al. (2013) are recent studies of the TV market. For healthcare markets, Gowrisankaran et al. (2015) and Ho and Lee (2015) are recent examples. Gaynor and Town (2011) surveys the earlier work.

use B and their value for B . For the analysis in Horn and Wolinsky (1988), it is sufficient to assume that the quantity for all other hospitals is unaffected. However, the empirical literature cannot make the same assumption and instead models the distribution of A 's patients among the other hospitals, as we do here. For example, Gowrisankaran et al. (2015), assume that all of the insurer patients would remain if a negotiation with their hospital breaks down.

If A leaves the network, the insurer pays B the same for a patient that would choose B over A and vice versa. Suppose, for example, that B 's per-patient price is set very close to the value per patient that would choose B . If A 's patients value B less (e.g. due to distance), then these patients create a negative surplus (value less price) without A in the network. As a result, A 's marginal value per patient at B 's equilibrium price may be higher than its actual per-patient value.

This dynamic guarantees that there is always a cutoff $\bar{\beta} < 1$ such that for every hospital bargaining power $\beta \geq \bar{\beta}$ the prices paid by the insurer are higher than the value of the service to the patients. That is, the insurer creates negative surplus with A in the network. However, the insurer creates even less surplus off-equilibrium without A in the network. The surplus maximizing choice of the insurer is to quit the market, resulting in the insurance market breakdown.²

This market breakdown results from four important features of the market. First, the hospital-insurer bargaining is mostly based on fee-per-service. Alternative contracts, e.g. lump sum payments as in Gowrisankaran et al. (2015), solve this problem, but give rise to other issues. Second, consumers pay only a small fraction of the price of hospital services they use, and instead pay a “subscription” fee – the insurance premium. This implies that consumers may use a service that is surplus negative when accounting for the full cost. Third, the consumer's choice of insurers cannot guarantee that the consumer will switch insurers to have ex-post access to their most favored hospital in case the bargaining ends with some hospitals not in the insurer's network. Fourth, insurers often consider the surplus they generate to their enrollees rather than simply their current net revenue.

²Health insurers in effect provide two distinct services to their enrollees – insurance and price bargaining. Our analysis abstracts away from the insurance provision service and considers only the implication of the price bargaining service. The insurer may create surplus by providing actual insurance. However, roughly half of the privately insured Americans are part of insurance plans in which the insurer provides all the services *except* the actual insurance, while their employer absorbs the risk (i.e. insures the enrollees). We expand on this in section 3.1

These features are common to the current empirical assumptions for the US healthcare market. The recent handbook chapter by Gaynor and Town (2011) (henceforth GT) treats the first three assumptions as standard for the related empirical work. The literature is split with regards to the fourth assumption, namely whether it is more appropriate to model insurers as maximizers of short-term profit or enrollee surplus, with compelling arguments in favor and against both.³ Our proposed RSN model is appropriate for either surplus or profit maximizing insurers, while the novel insurer market breakdown in NiN applies only to insurers that maximize (a fraction of) enrollee surplus. We discuss the implications for markets with short-run profit maximizing insurers in detail in section 3.1.

Our proposed RSN model is based on the behavioral assumption that the insurer can commit (at least temporarily) not to reopen negotiations with a hospital. Our impression is that insurers have sufficient mechanisms to make such commitment. For example, a negotiating unit that is limited by capacity to only negotiate with a small fraction of the industry’s providers at a time would accomplish this.

To gain insight into this question, we interviewed David Klein, a former CEO of a Blue Cross and Blue Shield in upstate New York – Klein (2015). Klein confirmed that insurers lack the resources (negotiators, actuaries, accountants, etc.) to negotiate with all their hospitals at the same time. Moreover, as a negotiating tactic, insurers avoid simultaneous negotiations even if those were possible as those are likely to increase the hospitals’ negotiating power and ability to coordinate. Instead, the insurer seeks to negotiate with each hospital once every several years and generally do not reopen negotiations.⁴

³ Town and Vistnes (2001), Capps et al. (2003) and Lewis and Pflum (2015) assume enrollee-surplus maximizing insurers, in part arguing that this is a better proxy for the insurer’s long run profits. In contrast, Gal-Or (1997) and Ho (2009) assume short run profit maximization. Most recently, Gowrisankaran et al. (2015) estimate their model for both surplus and profit maximizing insurers but focus on the surplus maximizing specification.

⁴ Insurers may also simply commit to have a limited network. In the extreme case that the insurer bargains with a single hospital, the aforementioned bargaining interdependency problem obviously cannot arise. The resulting networks will be smaller than the surplus-maximizing one, but may nevertheless generate higher surplus. However, again, holding all other agreements fixed, the insurer increases the surplus it generates by adding hospitals to its network. However, this is inconsistent with the observed networks in reality. Moreover, Wedig (2013) finds that consumers insurance choice is consistent with significant welfare costs for having access to a smaller hospital network. We view commitment to an order of negotiations as weaker than an a-priori commitment to a small network. This weaker form of commitment is enough for delivering our good news: sequential negotiations results in positive surplus and lowers average prices.

To construct the RSN model we first replace the simultaneous Nash approach with sequential bargaining. In RSN the insurer negotiates with the hospitals one-by-one. Once bargaining between an insurer and a provider concludes, renegotiation is impossible for some time. We then place this bargaining protocol within a repeated interaction framework that allows renegotiation in the future. Other than the bargaining protocol, the RSN model therefore can be implemented with any of the other assumptions currently used in the literature for the NiN model.

Extending the NiN estimation in Gowrisankaran et al. (2015) to RSN requires adding one parameter (the discount factor) and modifying the structural equation used to recover costs and bargaining parameters. RSN estimation can be viewed as a generalization of NiN that does not increase the complexity of the estimation procedure. As the estimate of the discount factor decreases, estimation based on the RSN model approaches the NiN model.⁵

In applications, the bargaining model is used to estimate the share of the surplus that is allocated to the insurer and the unobserved hospital costs. Section 5.2 shows that using the NiN and RSN models provide different estimates. In particular, estimates of hospital bargaining power and costs will be biased downwards when using the NiN model if the true underlying model is RSN. This is consistent with the finding in Gowrisankaran et al. (2015), where for two of the four insurers, the estimation obtained exactly zero hospital bargaining power or an extremely low hospital treatment cost under the assumption of NiN bargaining.

The bargaining model directly relates to the important question of merger evaluation. Section 5.4 shows that extending the model from NiN to RSN can qualitatively reverse the predictions and counterfactuals. In particular, mergers that the NiN model suggests would decrease prices may actually increase prices according to the RSN model. In particular, the NiN model would suggest that a merger between two hospitals that are high priced and relatively substitutes may actually decrease prices due to the elimination of the dynamic identified in section 2. In contrast, the RSN model would predict that such mergers will increase prices. Section 5.4 shows this qualitative result applies also if insurers are short-term profit maximizers.

⁵While the estimation (conditional on prices) is generalized, the RSN model with myopic hospitals does not converge to NiN. In particular if hospitals are completely myopic, the RSN model does not predict unique equilibrium prices.

While our exposition uses the hospital-insurer bargaining context, the analysis is also relevant to other important settings. In particular, large organizations (firms, hospitals, military, etc.) often have dedicated “purchasing departments” that negotiate with multiple “approved vendors”, while the actual product choice is done by professionals (e.g. doctors, engineers, military commanders) based on the products’ features on a case-by-case basis. Such situations map directly to our model. The objective of the “purchasing department” is to maximize the surplus generated to the firm from the acquisitions.⁶ In these settings the eventual customer is either unable or unlikely to change an intermediary when a specific supplier quits the network and the intermediary’s success relies on maximizing the value for its customers.

Media markets, in particular TV and music streaming are also closely related markets. Crawford and Yurukoglu (2012) and Lee et al. (2013) apply a similar framework to the TV market mentioned above and the RSN model can be applied there as well. However, in contrast with healthcare markets, in the TV industry currently the supplier payment is in general based on the distributor (“insurer”) revenue from bundles that include the supplier rather than actual consumption. Interestingly, the music streaming industry implements a different model in which supplier payment is based on the supplier’s ex-post share of the consumed content.

The next Section reviews the related bargaining literature in general, and in the specific health economics aspect; Section 2 provides the NiN result in a simple model that is generalized and analyzed in Section 3. Section 4 develops the RSN model. Section 5 develops the results for estimation and inference. Section 6 concludes.

1.1 Literature Review

In terms of the theoretical framework, our paper belongs to a strand of literature that concerns bilateral bargaining in vertically-structured markets.⁷ In this regard, the main reference to our work is that of Horn and Wolinsky (1988), who apply the

⁶Hospitals often form a Group Purchasing Organization (GPO) which negotiates on behalf of multiple hospitals with the same supplier. GPOs have been studied extensively, mostly in the supply chain literature. See Cleverley and Nutt (1984), Schneller (2000) and Scanlon (2002) for reviews of GPO effectiveness.

⁷A more general framework is that of bargaining in networks. (See e.g., Abreu and Manea, 2012; De Fontenay and Gans, 2013; Manea, 2011; Stole and Zwiebel, 1996).

asymmetric Nash bargaining solution to determine input prices in a duopoly model in which two firms acquire inputs from a common upstream supplier. More recently, Iozzi and Valletti (2014) considered a similar setting, in which each of N identical downstream firms buys its input from a single supplier at a first stage, which is followed by competition (among the firms who made a non-zero input purchase) at a second stage. As in Horn and Wolinsky (1988), Iozzi and Valletti (2014) assume that each of the N input prices is determined through bilateral Nash bargaining. Iozzi and Valletti (2014) consider the case where the second-stage competition is either a Bertrand competition or Cournot competition.

Binmore et al. (1986) and Davidson (1988) provide extensive-form models that result in the NiN solution, but are based on different underlying assumptions than those we study here. Collard-Wexler et al. (2014) provide a similar model motivated by recent applications. However, Collard-Wexler et al. (2014) depart from the empirical literature by assuming that insurers and hospitals negotiate a lump sum payment rather than fee-per-service also guarantees positive surplus, though raises additional concerns. Our critique of the NiN approach is generically independent of the underlying extensive form game that is assumed to generate the equilibrium.

The central difference between our work and the aforementioned papers is in the bargaining stage.⁸ Both in Horn and Wolinsky (1988) and Iozzi and Valletti (2014), the upstream-downstream negotiations are modeled as a collection of (simultaneous) independent bargaining problems. In our model, by contrast, there are significant interdependencies between the problems. Indeed, the main contribution of our analysis is the realization that the interdependencies are so strong that even the most basic intuitive properties of a bargaining solution may be violated. Crucially, the objective of the upstream entity of our model—the insurer, or, to use a more general term, the *collective bargainer*—is not a linear revenue-function; instead, the collective bargainer’s payoff (the expected welfare of the mass population it represents) is sensitive to the composition of the downstream market.

From the theoretical bargaining literature, the central work about interdependencies among bargaining problems is by Bennett (1997). She considers the case

⁸Iozzi and Valletti (2014) also distinguish between (1) the case where negotiation breakdown between the upstream monopolist and a particular firm i is observable by i ’s rivals and (2) the case where it is not observable. Our analysis does not rely on downstream (insurer) competition. However, if such competition exists, we follow the literature and assume that the negotiation outcome (i.e. insurer network structure) is common knowledge.

where $n \geq 3$ players form (possibly several) coalitions, bargaining takes place within each coalition, and the interdependencies among the different bargaining problems are reflected in the endogenous determination of the outside options: the disagreement payoff of player i in the bargaining problem in which he participates is calculated on the basis of what would have happened had that player belonged to another coalition (i.e., what payoff he would have obtained then). There are two important differences between Bennett’s work and ours. First, in her work each player can participate in at most one coalition. In our work, by contrast, the collective bargainers bargains with each hospital separately; that is, he simultaneously belongs to several coalitions. Secondly, at the conceptual level, one of the main issues in Bennett (1997) is that of coalition formation. In our work, by contrast, this is not an issue, due to exogenously given vertical structure: it is a priori clear what are the possible coalitions to look at, and, moreover, it is clear what is *the* coalition-configuration to focus on—the one corresponding to the full hospital-network.

Various studies consider sequential bargaining. Section 6 in Horn and Wolinsky (1988) considers the sequential offer equilibrium of the game there. Noe and Wang (2004) shows, in a general setting, the benefit of hiding the order of negotiations, a possibility we do not consider here. Gal-Or (1999) considers both simultaneous (NiN) and (one-shot) sequential bargaining between hospitals and insurers. However, the analysis there considers the insurer’s incentive to exclude a hospital from the network and the implications for industry consolidation.

Our approach of embedding the one-shot game in a repeated framework resembles that of Rey and Tirole (2013), who use the properties of the repeated game in a patent pool formation. The bargaining setting and potential inefficiencies there are however very different.

Our application of the Nash bargaining solution in the context of health economics closely follows the handbook chapter by Gaynor and Town (2011), which builds on numerous works from the health economics literature. Regarding the Nash bargaining that takes place between an insurance provider and a hospital, Gaynor and Town write (p. 530) “if the net surplus from the hospital-insurer match is not greater than zero, then bargaining does not take place and the hospital is not in the insurer’s network.” One informal way to interpret our negative-surplus result (Theorem 1 below) is that the aforementioned circumstances are actually plausible in many settings, if one accepts the simultaneous Nash model as appropriate.

2 The basic idea: two symmetric hospitals

A symmetric setting provides a simple illustration of the underlying dynamic resulting in higher-than-value prices. The next section generalizes the result.

An insurer negotiates with two hospitals, $j \in \{A, B\}$, serving a specific market. Wlog, normalize the cost of serving a patient to zero. The market has two types of patients, denoted $\theta \in \{a, b\}$. Assume for now that patients of type a (resp. b) have a value of 10 from being served by hospital A (resp. B) and 5 from being served by the other hospital. With both hospitals in the network, the insurer expects a unit mass of patients for each hospital. However, if hospital A leaves the network, only a fraction $\alpha \in (0, 1)$ of the a patients remain in the network.⁹ Similarly, if B leaves, α of the b patients remain. That is, the hospitals are exactly symmetric.

The insurer maximizes surplus to the patients (or a fraction of it). Let V denote the patient surplus of the full network and V^{-j} the same without hospital j . Negotiations are over prices p^j and set each hospital's bargaining power at $\beta \in (0, 1)$. Thus: $V = 20 - p^A - p^B$ and $V^{-A} = \alpha \cdot (5 - p^B) + (10 - p^B)$.

Under Nash bargaining, if we hold p^B fixed then p^A solves

$$\max_p (V - V^{-A})^{(1-\beta)} \cdot p^\beta \quad (2.1)$$

The price response is given by:

$$p^A = \beta \cdot [10(1 - \alpha) + \alpha(5 + p^B)] \quad (2.2)$$

Equation 2.2 shows that A obtains a fraction β of the surplus it generates. The $(1 - \alpha)$ new patients each account for 10 utils. The α patients that instead would have went to B only gain 5 directly from going to their preferred hospital, but also save the payment of p^B .

Of these two consumer segments $(1 - \alpha)$ and α , prices may surpass surplus only because of the second (α) group. In particular, whenever $p^B > 5$, the insurer actually generates negative surplus (net of prices) serving these patients without A in the network. The surplus that A generates to these patients is then higher than it's ex-post per-patient value of 10. If the hospitals' bargaining power is sufficiently

⁹allowing for some of the other type of patients to leave has no implication for the formal results in equation 2.4 or proposition 1 below but complicates the exposition.

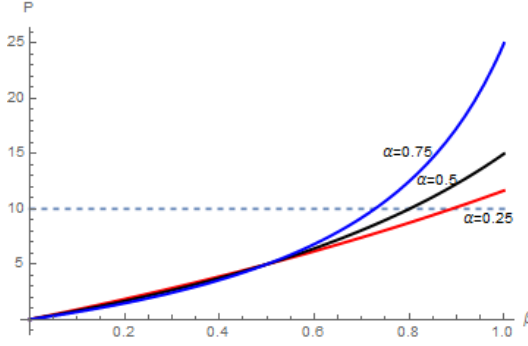


Figure 2.1: Symmetric example for $\alpha = 0.25, 0.5$ or 0.75

high so that hospitals obtain most of the surplus they generate, price will then be higher than the ex-post value.

Formally, prices are obtained by solving 2.2 and the symmetric equation for p^B . This obtains:

$$p^j = 5\beta \frac{2 - \alpha}{1 - \beta\alpha}, \quad (2.3)$$

for both $j \in \{A, B\}$. Recalling that with both hospitals in the network the value of each service is 10, the following is easy to verify:

$$p^j \leq 10 \iff \beta \leq \frac{2}{2 + \alpha} \quad (2.4)$$

That is, *as long as some patients stay with the insurer despite ex-post preferring a hospital that opts out of the network (i.e. $\alpha > 0$), there is some lower bound on the bargaining power parameter, $\bar{\beta} < 1$, such that for any $\beta > \bar{\beta}$, the surplus generated by the insurer is negative – each service's price is higher than its value.*

Figure 2.1 illustrates the resulting price per patient. Even if just a quarter of the patients stay with the insurer even if their favored hospital leaves the network, prices exceed value when $\beta \geq 0.88$. As the share of patients that would not adjust their insurer choice increases, equilibrium prices significantly increase and may well be more than double the value per patient.

The example is easily generalized. Let the *2-hospital symmetric model* be defined as follows: v is the value of a hospital to the patients that would go to it if both hospitals were in the network (i.e., 10 in the example) and $\lambda \cdot v$ be the value per patient of going to their second best hospital (i.e., $\lambda = 0.5$ in the example). Then again derive V , V^{-j} and the bargaining solution for each hospital as in (2.2)

to obtain:

Proposition 1. *In the 2-hospital symmetric model, the insurer generates negative surplus per patient (i.e., $p^j > v$) iff*

$$\beta \geq \frac{1}{1 + \alpha(1 - \lambda)} \quad . \quad (2.5)$$

In particular, for any valid parametrization $(\alpha, \lambda \in (0, 1))$, there is some $\bar{\beta} < 1$ such that for any $\beta > \bar{\beta}$ the price of each service is higher than its value

3 A general model

This section uses a variation on the handbook model of Gaynor and Town (2011) (GT) to generalize the violation of the PSP property. The next section then provides the alternative, repeated-sequential model that recovers PSP. To simplify the mapping between our result and the GT model, we fully adopt their notation, but slightly simplify the model. GT’s model allows for multiple insurers and hospitals, all potentially differentiated in costs and value of service. The main qualitative difference in our model relative to GT is that we assume insurance maximize their insured surplus rather than profit. As we reviewed in Section 1.1, the literature seems split on which is correct. However, an important implication of the current analysis is that this distinction has greater implications than perhaps was previously realized (see e.g., footnote 22 in GT).

Our model takes on the cooperative approach to bargaining. Namely, we do not specify a concrete extensive form, but only describe the economy in terms of its payoff-relevant information. We assume (as in the example from the previous Section) that prices are set as to maximize the relevant Nash products. In terms of the terminology and phrasing that we use in the text, we allow ourselves a little degree of informality, which manifests itself in two ways: (1) when we write “in equilibrium...” we do not refer to an equilibrium of any specific game, but simply to the situation in which all the prices in our model are set at their Nash-product-maximizing values; (2) when we write descriptive statements such as “when hospital j leaves the network,” or “the insurer proposes a price to the hospital,” this is only meant for the sake of illustration—it is not a description of a move in some extensive form.

The model is as follows. There is a measure of $\frac{1}{\rho}$ patients, of which a fraction of ρ will need hospital service. There is a fixed (finite) set J of (“candidate”) hospitals. A *network* is a non-empty subset of J . A generic network will be written as J_h (following GT, the index h stands for the insurer, and J_h is the insurer’s network). Should patient i need hospital service, she obtains utility $u_{ij} = f(x_j, z_i, d_{ij}) + e_{ij}$ from using hospital j , where z_i and x_j are, respectively, patient and hospital specific parameters, d_{ij} captures joint parameters (e.g., distance from patient to hospital) and e_{ij} is an unobserved logit error term (i.e., type-1 EV) that is only revealed to the patient when requiring hospital service. The patient can also choose an out-of-network hospital, which provides zero utility by assumption.¹⁰ Compared to GT, the simplification here is that we bundle all possible diagnosis to one (instead of M possible diagnosis in GT). This has no qualitative implications.

The logit assumption implies that the ex-ante value of a network J_h to a patient is:

$$W_i^{J_h} = \ln \left(1 + \sum_{j \in J_h} \exp(f(x_j, z_i, d_{ij})) \right)$$

In the GT framework, each insurer h agrees with each hospital in its network on a fee per patient service p_{jh} and then sets premiums knowing all networks and assuming the standard logit model for patient insurance choice with W_i as patient utility and some premium sensitivity parameter α . This results in insurer gross revenue $F_h(J_h)$ and hospital ex-post quantity demanded from the insurer q_{jh} . The same can be calculated for the network without any single hospital j , which will be denoted as J_{h-j} . Let c_j denote the hospital’s cost per patient-service, r_j the hospital’s outside option and cm_{jh} some hospital-insurer specific fixed costs. Then the Nash bargaining solution provides equation (9.7) in GT as the equilibrium price-response for hospital j from insurer h (the equivalent of equation 2.2). We provide it here verbatim, multiplying both sides by q_{jh} for clarity.

$$p_{jh}q_{jh} = (1-\beta)(q_{jh}c_j - r_j) + \beta \left[F_h(J_h) - F_h(J_{h-j}) - cm_{jh} + \sum_{l \neq j} p_{lh} \cdot (q_{lh}^{J_{h-j}} - q_{lh}) \right] \quad (3.1)$$

The hospital’s total revenue is a bargaining-power weighted average of its costs

¹⁰Both the logit assumption and the out-of-network assumption are made in GT and used here to simplify the exposition. Our model is agnostic about these. The formal assumptions required for the result are explicitly identified below.

and the net benefit to the insurer, holding all other prices fixed. The difference $F_h(J_h) - F_h(J_{h-j})$ is the increase in revenue for the insurer. The sum at the end is, as in our simpler model from the previous Section, the change in payments to other hospitals l due to patients that go to these hospitals without j in the network but otherwise would not. Both r_j and cm_{jh} are exogenous variables that do not contribute to the theoretical analysis. We therefore simply set these to zero.

In addition to allowing for multiple hospitals and insurers, the GT model assumes insurers maximize revenue both in their network choice and their pricing choice. Whether insurers maximize profits or surplus (net of hospital payments) is an open empirical question. We assume for the main analysis that insurers are surplus-maximizing. Therefore, we replace the F_h interpretation in 3.1 with the expected value for the patients. This allows us to abstract away from the patient's decision to pay the insurer premium. We discuss profit-maximizing insurers separately below. To evaluate surplus, we let $w_{h,j}$ denote hospital j 's average value to h 's patients that choose j in equilibrium (see equations 9.2 and 9.3 in GT).

As in our simple 2-hospital benchmark, the key assumption in the analysis is that some patients remain with an insurer even if their ex-post preferred hospital leaves the network. These patients will now use a different hospital, implying substitution between hospitals. Denote by $w_{h,j}^l$ the average value from j for h 's patients that choose l in the equilibrium network but would choose j if l is not in the network. To capture the notion of “second best,” we assume that patients generally value their second-best hospital less than the patients that freely choose it. That is:

Assumption 3.1. For any hospital network, for all insurers:

1. Hospitals are substitutes: $q_{lh}^{J_{h-j}} \geq q_{lh}$ with the inequality strict at least for one l for every insurer-hospital pair h, j
2. Second best values are lower: $\forall j, l : w_{h,j} > w_{h,j}^l$

We make an additional “independence of irrelevant alternatives” (henceforth, IIA) assumption that greatly simplifies the exposition. While stronger than required, the assumption is natural and would otherwise be replaced by cumbersome restrictions on patient preferences:

Assumption 3.2. IIA: If hospital j leaves a network, the hospital choice of patients that would have went to hospital $l \neq j$ with j in the network does not change.

Call the above model the *general model*. The main result for this Section is this:

Theorem 1. *In the general model, there is a $\bar{\beta} < 1$ such that for any $\beta > \bar{\beta}$ the surplus generated by each insurer is negative.*

Proof. For simplicity, we omit the h subscript. Let $x_l^j \equiv q_l^{-j} - q_l$ denote the quantity of j 's patients that move to l if j leaves the network.

Using 3.1:

$$p_j q_j = (1 - \beta) q_{jh} c_j + \beta \left(F - F_{-j} + \sum_{l \neq j} p_l \cdot x_l^j \right) \quad (3.2)$$

By construction, $F = \sum_l q_l \cdot w_l$ and $F_{-j} = \sum_{l \neq j} (q_l \cdot w_l + x_l^j \cdot w_l^j)$. This simplifies (3.2) to

$$p_j q_j = (1 - \beta) q_{jh} c_j + \beta \left(q_j w_j + \sum_{l \neq j} x_l^j \cdot (p_l - w_l^j) \right) \quad (3.3)$$

What we have here is $|J|$ equations in $|J|$ unknowns, $\{p_j\}_{j \in J}$. If all $x_l^j = 0$, then as $\beta \rightarrow 1$ we have that p_l is arbitrarily close to $w_l > w_l^j$. Now increase all of x_l^1 to their true value. As $p_l > w_l^1$, this strictly increases p_1 so that $p_1 > w_1 > w_1^j$ without affecting any other prices. Repeating the procedure for every j completes the proof. \square

3.1 Discussion

Theorem 1 applies to insurers whose objective is to maximize their enrollee surplus. In contrast, insurers maximize only total premium less total costs are modeled to not “care” if patients will use a hospital that provides less value than reflected in the premium, as long as the patients pay the premium. This makes the short-term-profit maximizing insurers immune from the negotiation tactic we described above.

As presented in the introduction, whether the network formation decisions of most insurers are better modeled as maximizing enrollee-surplus or short-term net profits is an open empirical question. Town and Vistnes (2001) suggest it is appropriate to model even profit maximizing insurers as surplus maximizers because, in the long term, patient choice reacts to surplus and as a result, an insurer's long term profit is ultimately best approximated as a fraction of the surplus it generates.

The various potential differences between profit and surplus maximizing insurers are outside the scope of the current paper. For our result to hold, the only substantial requirement is that insurers internalize the reduction in their enrollees welfare from not having access to the hospital of their choice. This could be through direct surplus accounting, as we model, or through concerns from churn or patient backlash.

An additional reason¹¹ to consider surplus maximization is that about one in two Americans covered in the private market are covered by a self-insured employer.¹² In those plans, the insurer¹³ negotiates prices with providers but does not collect the enrollee premiums. Instead, the insurer is typically paid based on costs or volume of claims and enrollees. We are not aware of a model of insurer competition in the self-insured market. However, as employers seem to consider health benefits a service to employees rather than a profit generator, surplus maximization seems a more appropriate proxy than profit maximization, with insurers competing in their ability to create value for the employer's plan.

One interpretation of our analysis is that under the conditions above, if bargaining matches the simultaneous model, *profit maximizing insurers may well generate more surplus than surplus maximizing ones*. In particular, surplus maximizing insurers may generate negative surplus, while profit-maximizing insurers will generate positive surplus (subject to imperfections of consumer insurer choice).

This insight is also important for other settings to which our theory applies. To be specific, consider a purchasing department that contracts with various competing providers over inputs to be used by several downstream "user" departments (e.g., a university's purchasing department negotiating with several laptop makers). If the department is evaluated as a profit center based on some transfer price (or willingness to pay) from each of the downstream departments *and* the downstream users must make their purchases through the upstream purchasing department, then the organization will be subject to the same problem as the surplus maximizing insurer

¹¹We thank Henry Mak for suggesting this.

¹²Smith and Medalia (2014) show (table 1) that in 2013 64.2% of the US population was covered by a private plan, with 53.9% employment based and 11% direct-purchase, implying that roughly 84% of the population covered by private insurance was covered by an employment based plan. Of these, Agency for Healthcare Research and Quality (2015) report that roughly 58.2% are in self-insured plans. 58.2% of 84% is 48.9%.

¹³Insurers are more accurately referred to as Third Party Administrators (TPAs) when acting in this capacity.

and may well generate negative surplus. We expand on this in follow-up work.

A second implication is that if insurers are surplus maximizing and simultaneous Nash applies, it may be more efficient to let hospitals merge. If all hospitals in a market merged, the merged hospital will provide each patient the best option. The insurer is never threatened with negative patient surplus and as a result, the merged hospital can never obtain more than the entire ex-post surplus it generates. A direct implication of Theorem 1 is that for sufficiently high β , this is the only market structure that protects a surplus-maximizing insurer from generating negative surplus.

Corollary 1. *In the general model, there is a $\bar{\beta} < 1$ such that for any $\beta > \bar{\beta}$, surplus from a enrollee-surplus-maximizing insurer is maximized if and only if all hospitals merge.*

We conclude the discussion with noting aspects of consumer choice in insurance markets that raise additional issues that do not directly impact our analysis vis a vis the GT model.

First, patients can react to a breakdown in insurer-hospital negotiations by *switching* to the insurer network that provides them the most value conditional on their predictions and premiums. Handel (2013) found that inertia (or switching costs) plays a significant role in consumer insurance choice. This implies also that each hospital contributes very little to the insurance revenue and prices. As a result, hospital margins and prices will be low.¹⁴

The opposite concern regarding switching is that patients could be very risk averse and switch out of smaller networks. Wedig (2013) found that such concerns made smaller networks unprofitable for insurers. At the extreme, smaller networks are never chosen by consumers and the net value of each hospital to the insurer in the simultaneous bargaining approach (as in GT) is the entire revenue per patient, leading to a market breakdown.

Finally, Sorek (2015) shows that accounting for patient uncertainty about their preferred provider at the time of insurance selection has significant implications for the resulting equilibrium product choice and price.

Incorporating a more detailed consumer choice model for insurance selection therefore is an important avenue for further research.

¹⁴The surplus implications of such high switching costs can be significant, both in terms of high insurance premiums and suboptimal insurance choice by consumers.

4 Repeated Sequential Nash

4.1 Sequential Nash

The root of the overpricing which gives rise to the negative surplus from Theorem 1 lies in the hypothetical “disagreement event” associated with each negotiation. For example, in the two-hospital case, when A is out of the network (i.e., there is disagreement with A) its entry-contribution exceeds its true value because its absence from the market would generate an adverse shift in patients going to hospital B which would be welfare reducing at B ’s equilibrium price. Avoiding this counter-intuitive outcome crucially depends on constructing a bargaining mechanism where the impact of disagreement on other bargaining outcomes is a part of the analysis.

One way to achieve lower prices is to assume that once there is a negotiation breakdown with some hospital, other hospitals observe the breakdown and negotiation continues accordingly. This can be achieved by sequential bargaining: the hospitals are ordered in a (commonly known) sequence. If negotiation breaks down with some hospital early in that sequence all subsequent negotiations assume the rejecting hospital is not in the insurer’s network. Such sequential bargaining have been frequently considered as a possibility in theoretical work, including section 6 in Horn and Wolinsky (1988).

It is easy to see that the insurer’s surplus is positive under sequential negotiations. Consider again the two-hospital case with A going first and B going second. Given any outcome of negotiations with A , the bargaining problem with B must result in a non-negative addition to the insurer’s overall surplus (or else the insurer will not sign a deal with B). Also, one can map any possible outcome in the negotiations with A , say o , to the subsequent bargaining problem that will be played with B , say $P(o)$. Since, as we just observed, the insurer’s surplus in $P(o)$ is non-negative given any possible o , the bargaining problem with A boils down to a standard Nash bargaining problem in which both parties make positive profits. This idea generalizes to any length of hospital-sequence.

The order of negotiations affects hospital payoffs. For example, in the two-hospital case where A is the first in the sequence, disagreement with A automatically makes B a monopolist. In contrast, disagreement with B cannot have such a favorable effect on A ’s bargaining position, since it can only happen after the interaction with A has concluded; specifically, either (i) a deal with A has already been

signed and so A 's price is fixed, or (ii) A has dropped out, and is no longer in the network.

In terms of formalism, the economic environment that we consider is the same as in the previous Section, with the only difference being that the Nash products now reflect the order of negotiations. Also, we add a technical assumption about off-equilibrium beliefs. We assume that the insurer and hospitals assume that if two hospitals leave the network then the patients that had those two hospitals as their top two choices, also leave the network. This allows us to ignore “high-order” bargaining effects¹⁵ that would be negligible when β is sufficiently large in any case. Technically, this keeps the steps of the proof from growing linearly in the number of hospitals. We note that the assumption is vacuous in a two-hospital setting. Call this the *sequential model*.

The positive surplus result (for the insurer) and the favorable status of late bargaining positions (for the hospitals) are summarized in the following results.

Proposition 2. *Assume the sequential model. For any $\beta \in (0, 1)$, the surplus generated by each insurer is positive. Moreover, the insurer's surplus is independent of the order of negotiations with the hospitals.*

Appendix A.3 provides the closed form solution for the sequential model – prices per hospital in the insurer's expected value on which the propositions are based.

While the insurer is indifferent about the order of negotiations, the hospitals are not. Proposition 3 confirms the intuition above and shows that in a two-hospital setting or if all hospitals are symmetric as in section 2, being later in the order increases the hospital's profits.

Proposition 3. *Assume the sequential model with two hospitals or with J symmetric hospitals. If $\beta \rightarrow 1$, negotiating later in the order increases a hospital's price and profits.*

¹⁵Incidentally, in our interview, Klein (2015) without this assumption being mentioned at all, indicated that as an insurer, he expected backlash but not significant loss of business if one hospital left the network, but a drastic reaction and loss of business if two (or more) hospitals leave. This is exactly the content of the assumption. That is, the hospitals and insurers assume that if A leaves the network, each patient that had A as its preferred hospital and remains in the insurer's network would leave the network if their second option hospital also leaves the network.

As prices are complementary, the price paid to earlier negotiators also raises the price to other hospitals. In other words, the marginal cost of each price for the insurer is higher early in the negotiation sequence. As a result, negotiating early in the sequence reduces the hospital's price per patient. As the hospitals' quantity is independent of the negotiation order, the price effect determines the profit effect. It is difficult to derive more general comparative statics for hospital profits based on order. However, as hospital prices are strategic complements and hospitals are substitutes, the underlying economics imply that going first will be worst for any hospital than going last or obtaining the NiN price. This result is similar to proposition 3 in Horn and Wolinsky (1988), who also find that unless the insurer (supplier there) has a value for failing both negotiations that exceeds some strictly positive threshold, the second negotiator obtains better terms. As the insurer's outside value is zero here (and in all related studies), the implication of Horn and Wolinsky (1988) is that the worse position is first.

4.2 Repeated Sequential Nash

We now extend the single sequential interaction to an infinite horizon setting, in which the aforementioned interaction—namely, the sequential model—takes place in each and every period. Call this the *repeated sequential nash* or RSN.

We adopt the Nash Reversion model where in each period, each insurer approaches each hospital with an offer. If all hospitals accept the insurer's offers prices are set for the period. If a hospital counters the insurer's offer, this is considered a deviation. We assume that once a hospital deviates in a given period, the Nash bargaining result is obtained between the countering hospital and the insurer for this period, taking all other prices at the equilibrium level as usual. In all periods after a deviation, the insurer reverts to the one-shot Sequential Nash with a known order, in which the hospital that countered is placed first.

The previous section established that first is the worst position in a two hospital setting and with symmetric hospitals. While there could be worse punishments, using this provides a unique and well defined punishment for all applications. More generally, the methods in Abreu (1986) can obtain lower punishment prices by requiring lower pricing by the other hospitals in the punishment phase. However, as we cannot motivate these empirically, we simply note that these would be analogous

to assuming a lower hospital bargaining power.

As before, since a detailed description of the extensive form will not be needed in the sequel, we specify only payoff-relevant data. Within each given period, payoffs are as above. The only difference is that now each agent (insurer, hospitals) seeks to maximize the discounted sum of its period payoffs from each time period onwards.

We assume all hospitals have a common discount factor $\delta \in (0, 1)$.

The following notation will be used in the remainder of the paper:

- F_{jh} is the value of insurer h 's patient surplus (or expected revenues) with *only* hospital j in the network.
- F_h is, as before, the value of the insurer's patient surplus (or expected revenues) with the equilibrium network.
- W_j is the increase in the insurer's objective from adding hospital j to the network, given all other hospitals' prices (i.e. the bracketed term in any of 3.1, 3.2 or 3.3)
- v_j is the expected value per patient treated for hospital j given the observed network.
- $v_{j,i}$ is the expected value for a patient that would go to hospital j if it was in the network but instead goes to hospital $i \neq j$ if j is not in the network.
- $q_{j,i}$ is the number of patients that would go to hospital j if it was in the network but instead go to hospital $i \neq j$ if j is not in the network.
- c_j is j 's unobserved marginal cost and c_{-j} is a vector of $J - 1$ length of j 's rivals' (unobserved) costs.
- p_j is hospital j 's equilibrium price.
- \hat{p}_j is the price the hospital would obtain if the insurer reverted to the sequential Nash and ordered the negotiation with j first.
- q_j is hospital j 's expected number of patients in the full network.
- q_j^{jh} is hospital j 's expected number of patients if it is the only hospital in h 's network.

We show in appendix A.3 how to derive the punishment price \hat{p}_j . While tedious, this is a purely technical exercise.

Lemma 1. *In the RSN model:*

$$\hat{p}_j = \beta \frac{q_j v_j + \sum_{k \in J, k \neq j} [\beta q_{k,j} v_{k,j} - (1 - \beta) q_{j,k} (v_{j,k} - c_k)]}{q_j + \beta \sum_{k \in J, k \neq j} q_{k,j}} + (1 - \beta) c_j \quad (4.1)$$

Proof. Set $i = 1$ in equation A.11. □

The elements in the numerator multiplying the various v values capture the patient value that hospital j generates in equilibrium, plus β times the value j helps the insurer as an alternative to another hospital in the negotiation, less $(1 - \beta)$ times the value that the insurer will maintain without j thanks to other hospitals. Multiplying the first quantity by β and the second by $(1 - \beta)$ accounts for the fact that the insurer only enjoys surplus that does not get bargained away by any other hospital. In the extreme case that $\beta = 1$, hospital j captures all the surplus it generates as a second-best hospital and the insurer captures none of the surplus generated by the other hospitals as second-best alternatives to j . The elements multiplying rival costs capture the additional cost implication of j when it is a second-best hospital. The denominator normalizes the per-patient surplus generated by j to account also for the extent that j captures any value as second best.

An important qualitative result is that as $\beta \rightarrow 1$, any surplus generated from hospitals joining the network after the first hospital is captured by those later joiners. As a result, the price for the first hospital reflects only its' surplus as the single hospital in the network.

Proposition 4. *In the repeated sequential model, if hospitals have sufficiently large bargaining power the punishment price converges to the hospital's per patient value if it was the only hospital in the network: For every $\varepsilon > 0$ there is a $\bar{\beta} < 1$ such that for all $\beta > \bar{\beta}$, $|\hat{p}_j - \frac{F_{jh}}{q_j}| < \varepsilon$.*

By assumption 3.1, the per-patient value a hospital provides is lowest when it is the only hospital in the network. This is because removing a rival hospital j' from a network only adds to j some patients that would prefer j' . Then j 's per-patient value for these patients is lower than for patients that choose j regardless of j' . Thus, starting from the full network and iteratively removing hospitals until only j

remains in the network decreases j 's per patient value at each iteration. Moreover, the “punished” hospital will obtain even less than its standalone value F_{jh} . This is because j 's equilibrium quantity q_j is smaller than its standalone quantity q_j^{jh} .

Corollary 2. *In the repeated sequential model, the profit for a “punished” hospital is strictly lower than the average per-patient value that the hospital provides.*

To determine the repeated equilibrium prices, we next consider the IC constraints for each hospital. If hospital j rejects the price proposed by the insurer, the bargaining that commences is over j 's added surplus W_j , holding all other prices fixed. This results in a deviation period payoff for j of $\beta \cdot W_j$, and so a deviation price of

$$p_j^D = \beta \cdot \frac{W_j}{q_j} + (1 - \beta)c_j. \quad (4.2)$$

Note that p_j^D is the same price that would be obtained for j in a NiN model analysis if all other prices were the observed equilibrium price. As this is a “Nash Reversion” equilibrium as described above, the price in all periods after deviation is the punishment price \hat{p}_j . This gives rise to the following IC constraint for hospital j :

$$p_j^D \cdot (1 - \delta) + \delta \hat{p}_j \leq p_j. \quad (4.3)$$

As surplus from the network is independent of prices, the insurer minimizes prices p_j subject to these J constraints. As W_j and the insurer's objective are linear in prices, all IC must bind. The solution of this problem provides the equilibrium prices of the RSN model:

Proposition 5. *In the equilibrium of the repeated sequential model, hospital prices are given by equations 4.2 and 4.3 as an equality.*

As usual, the equilibrium of the repeated game is sustained only for sufficiently patient agents ($\delta > \bar{\delta}$) for some $\bar{\delta} < 1$. If the agents are impatient, $\delta < \bar{\delta}$, only the NiN prices satisfy the repeated game constraints. This is because each hospital takes all other hospitals prices as fixed and acts “as if” it is last. In this case, the insurer simply reverts to any arbitrary one-shot sequential negotiation. This guarantees that the market never breaks. The insurer's total payments never exceed the total value of the services provided. In our example below, $\bar{\delta}$ is less than 0.3. Generally, as

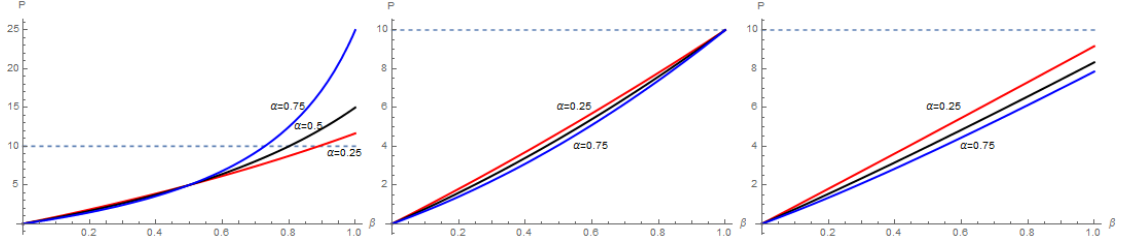


Figure 4.1: Prices for the symmetric example from section 2 for $\alpha = 0.25, 0.5$ or 0.75 . The left panel is for the 2-hospital symmetric model (as in section 2). The center panel shows the average price paid by the insurer in the sequential model. The right panel shows the limit of the price as $\delta \rightarrow 1$ for the repeated sequential model.

the NiN prices are the highest possible for the repeated equilibrium, the repeated equilibrium is sustained whenever the NiN prices do not imply market breakdown.

If hospitals are very patient, prices converge to the “punishment” levels. Since punishment levels provide each hospital with less profit than the value it generates (lemma 2), the insurer is guaranteed a positive surplus (or profit).

Proposition 6. *In the RSN model, if hospitals are sufficiently patient, the insurer’s per-period surplus (or profit) is strictly positive. If hospitals have sufficiently large bargaining power, the insurer’s profit converges to $F_h - \sum_j \left(F_{jh} \cdot \frac{q_j}{q_j^h} \right)$.*

Figure 4.1 compares the prices for the symmetric setting used as the motivating example in section 2 under three bargaining models: the standard model, the one-shot sequential bargaining, and the repeated sequential bargaining for patient hospitals. Moving from the standard to the sequential model restores positive surplus. However, as $\beta \rightarrow 1$, the hospitals are able to capture all the surplus. In particular, the second hospital in the negotiation obtains prices that are higher than its value, while the first hospital obtains prices that are lower. Moving to the repeated sequential bargaining model shifts the prices for both hospitals to the “first hospital” price, leaving the insurer a strictly positive margin even when the hospitals have all the formal bargaining power ($\beta \rightarrow 1$).

5 Implications for Empirical Work

This section generalizes the existing estimation equations to allow for both RSN and NiN bargaining. This is followed by an analysis of the implications for inference and counterfactuals. In particular for welfare and merger evaluation. We use the main specification of Gowrisankaran et al. (2015) (hereafter GNT) as the basis. The same steps can be done for the alternative specification in which the insurer maximizes its short term expected revenue.

5.1 Estimation

GNT first implement option-demand estimation as in Capps et al. (2003) to estimate the patients' utility for each hospital and each potential insurer network. This estimation is independent of the bargaining model assumed and can be used for our model as well.

We add two variables to the notation specified in section 4.2. First, for convenience of mapping the result to GNT, we follow their approach and add a parameter τ so that the insurance's objective function is

$$V(J) = \sum_{j \in J} q_j (\tau v_j - p_j)$$

In the standard model $\tau = 1$. Insurers that value their patients' utility more (resp. less) than the cost to generate the utility are described by $\tau > 1$ (resp. < 1). In the preferred specification in GNT, $\tau = 2.5$ with a standard error of 2.5. .

Second, we let p_j^G be the RHS from equation 14 in GNT. This is equation specifying the relationship between prices, the bargaining parameters, costs and estimated demand parameters. We replicate it here for convenience:

$$\text{GNT-14 : } p = c - (\Omega + \Lambda)^{-1} \cdot q$$

In GNT-14, the quantities are vectors of length J (i.e. one element per hospital).¹⁶ q is the observed demand and the two matrices Ω and Λ contain only various estimated demand values along with β and τ (nonlinearly).

¹⁶Here and in what follows, if a price appears without a subscript, the smallest unit in the equation is a vector rather than scalar.

The first step in the estimation process is to obtain an estimate of the punishment prices. Those are given in equation 4.1. To accommodate τ in the punishment price simply multiply all v values by τ there, obtaining:

$$\hat{p}_j = \beta\tau \frac{q_j v_j + \sum_{k \in J, k \neq j} [\beta q_{k,j} v_{k,j} - (1 - \beta) q_{j,k} v_{j,k}]}{q_j + \beta \sum_{k \in J, k \neq j} q_{k,j}} + \beta \frac{\sum_{k \in J, k \neq j} (1 - \beta) q_{j,k} c_k}{q_j + \beta \sum_{k \in J, k \neq j} q_{k,j}} + (1 - \beta) c_j \quad (5.1)$$

As in GNT, all the values in 4.3 except for β, τ and the hospital costs are derived from the demand estimation. Moreover, the key result in GNT for this step is maintained: the relationship between prices and costs is linear.

Proposition 5 shows that the observed price according to the RSN model is as in equation 5.2 if the deviation price is p_j^G . As noted there, equation 4.2 exactly implies this and therefore:

Corollary 3. *In the RSN model, estimated equilibrium prices satisfy:*

$$p_j = \delta \hat{p}_j + (1 - \delta) p_j^G \quad (5.2)$$

Corollary 3 and 5.1 allow us to derive the parallel equation to GNT-14. First, define θ as a vector that captures the first element in each \hat{p}_j :

$$\theta_j = \beta\tau \frac{q_j v_j + \sum_{k \in J, k \neq j} [\beta q_{k,j} v_{k,j} - (1 - \beta) q_{j,k} v_{j,k}]}{q_j + \beta \sum_{k \in J, k \neq j} q_{k,j}}.$$

Next, define Ψ as a matrix with off diagonal elements that capture the additional cost implication of j when it is a second-best hospital (by definition $q_{j,j} = 0$):

$$\Psi_{j,k} = \beta \frac{(1 - \beta) q_{j,k}}{q_j + \beta \sum_{k \in J, k \neq j} q_{k,j}}.$$

This provides a matrix representation for \hat{p}_j :

$$\hat{p} = \theta + (\Psi + (1 - \beta)I) c \quad (5.3)$$

Corollary 3, GNT-14 and equation 5.3 obtain:

$$\begin{aligned}
p &= \delta (\theta + (\Psi + (1 - \beta)I) c) + (1 - \delta) (c - (\Omega + \Lambda)^{-1} q) \\
&= \delta \theta + [\delta (\Psi + (1 - \beta)I) + (1 - \delta)I] c - (1 - \delta)(\Omega + \Lambda)^{-1} q \\
&= \delta \theta + ((1 - \delta\beta)I + \delta\Psi) c - (1 - \delta)(\Omega + \Lambda)^{-1} q
\end{aligned} \tag{5.4}$$

The first line in 5.4 places the corresponding values for \hat{p} and p^G . The second line collects the coefficients for c and the final line simplifies. The final step of the estimation in GNT consists of solving GNT-14 for c , obtaining GNT-19.¹⁷ The same is done here using equation 5.4 to obtain:

$$c = ((1 - \delta\beta)I + \delta\Psi)^{-1} \cdot (p - \delta\theta + (1 - \delta)(\Omega + \Lambda)^{-1} q) \tag{5.5}$$

As in GNT, equation 5.5 recovers the marginal costs for each hospital as a function of demand characteristics, β , τ and δ .

We can now state the generalization result.

Proposition 7. *Hospital costs implied by NiN model estimation are equivalent to those implied by RSN model estimation constrained to $\delta = 0$.*

Proof. Setting $\delta = 0$ in 5.5 obtains

$$c = p + (\Omega + \Delta)^{-1} q$$

as in GNT □

Proposition 7 confirms that the NiN model is a special case of the RSN model. In particular, if the estimated value for δ is zero, then the data supports the NiN interpretation, while a δ closer to one supports the repeated interaction model.

Estimation of the supply side parameters using RSN follows the procedure in GNT, replacing GNT-14 with 5.5. This implies estimating one additional parameter (δ), so that the total parameter set to be estimated is (β, δ, τ) and the cost parameters.

¹⁷GNT-19 has an additional term γv which is an extra cost fixed-effect term per hospital. This can be added directly to c here as well.

5.2 Implication for Inference

The NiN estimation is equivalent to RSN estimation constrained with $\delta = 0$. As $\delta \rightarrow 1$, estimates based on the NiN model may be inconsistent with the RSN model. As discussed in the previous section, the bargaining model affects the estimation of the hospital costs and the relative bargaining power.¹⁸ Therefore, the remainder of the section compares the main implications of using the NiN for inference when in the true model $\delta \rightarrow 1$.

For every demand parameters and observed rival prices, restricting $\delta = 0$ is equivalent to letting each hospital get its deviation profit, which is by construction higher than its repeated game profit. As quantities are fixed, this implies that the NiN model estimates larger margins for the hospital for every assumed β . That is, true margins for every β, c are lower than predicted by the NiN model.

Suppose first that β is perfectly observed (or assumed fixed) and that the true cost is c . Because the NiN model predicts a larger margin (i.e. price) for every cost, the NiN estimation will result in a lower c for the observed price. Alternatively, suppose that c is known and the true hospital bargaining power is unknown β . Because the NiN model predicts higher hospital profits than observed, the estimation will result in a lower β to justify the lower than predicted margins.

These two differences suggest that in practice both estimates will be too negative whenever restricting $\delta = 0$ if in fact $\delta > 0$. Indeed, GNT in table 5 specification 2 report that when β is estimated separately per each insurer the hospitals are facing, it is estimated at zero for two of the four insurers (insurers #2, 3 with S.E. 10^{-10}), 0.24 for another insurer (#4), and 0.5 for insurer #1 (both estimates with SE of 0.09).¹⁹ GNT's preferred model restricts $\beta = 0.5$ which implies that unrestricted, β is estimated lower than the preferred model by 100% in two cases and 50% in the third.

GNT table 5, specification 1 reports the results of the estimation restricting $\beta = 0.5$. As suggested above, the estimation obtains that the hospitals' cost of serving the patients of insurers #2, 3 is roughly 9,000 lower than for insurer #1 (S.E. ≈ 3000), whose cost is estimated at around 10,000. That is, a 90% decrease. The cost estimate for insurer #4 is roughly 4500 lower than average. Of course, without

¹⁸In GNT, the bargaining model also estimates their additional parameter τ . Our analysis here is restricted to the standard cost and bargaining parameters.

¹⁹The table there reports the insurer bargaining power, or $1 - \beta$.

re-estimating the model, we cannot determine whether these differences are a result of true cost differences across patients that are correlated with the patient's insurer choice or of the restriction $\delta = 0$.

5.3 Example - Logit Demand

A simple exercise with the Logit demand model illustrates the difference between the models and their predictions. Suppose there are n hospitals, each with the same base utility v and an outside good, whose utility to patients net of costs is u_0 . Assume for simplicity (as is done in GNT) that the patients rather than the insurer, fully internalize the cost of treatment from the outside good. Letting $S(n)$ denote the total market share of all goods in the market (i.e. one minus the outside good share), and $K \equiv \frac{n}{S(n)} \log \left(\frac{n+e^{u_0-v}}{n+e^{u_0-v}-1} \right)$ We show in the appendix that the NiN prices, p^N are given by

$$p^N = \frac{\beta}{1 - \beta \cdot S(n-1)} K \quad (5.6)$$

The Nash Bargaining solution implies complementarity in hospital prices, because it is assumed that in the out-of equilibrium case of disagreement between the insurer and one hospital, $S(n-1)$ of the hospital's patients will be served by the other hospitals at equilibrium prices will hold.²⁰ The case when the outside good is relatively weak (i.e. $u_0 \ll v$ and so $S(n-1) \rightarrow 1$), and hospital bargaining power is high ($\beta \rightarrow 1$) is particularly stark. The strong price-complementarity implied by NiN drives prices not just above insurer's surplus, but arbitrarily high.

Re-distribution of the disagreeing hospitals' patients to remaining hospitals, which by definition offer these patients lower utility, will affect other hospitals' average surplus and disagreement values. The key difference between the NiN and the RSN protocols is that only the latter allows for these effects to be taken into account by the bargaining parties.

In the RSN model, let p^R be the equilibrium price, p_1 the punishment price and p^D the deviation price. From proposition 5 we have that

$$p^R = (1 - \delta)p^D + \delta p_1 \quad (5.7)$$

²⁰ The NiN model in section 3 did not consider an outside good. If it would, then β in the denominator there would also be multiplied by the total shares of all inside goods in the market as here.

The deviation price for a hospital is the NiN price assuming all other hospitals price at p . The same procedure as for p^N above therefore results in

$$p^D = \beta(K + S(n-1)p^R) \quad (5.8)$$

Solving the two equations for the two unknowns p^D and p^R yields:

$$\begin{aligned} p^D &= \beta \frac{K + \delta \cdot S \cdot p_1}{1 - \beta \cdot S(n-1) \cdot (1 - \delta)} \\ p^R &= \frac{\beta K(1 - \delta) + \delta p_1}{1 - \beta \cdot S(n-1) \cdot (1 - \delta)} \end{aligned} \quad (5.9)$$

This illustrates that at the myopic limit (as $\delta \rightarrow 0$) the RSN model admits the same equilibrium as the NiN ($p^R = p^D = p^N$). Of course, as discussed in section 4, if firms are myopic, the insurer is better off using non-repeated sequential bargaining, which still guarantees positive surplus.

To illustrate the difference between the two models, set $u_0 = 0$ and $n = 11$ as in GNT. To simplify, assume that all hospitals have the same v and there are no additional random coefficients. Random coefficients represent an additional source of consumer heterogeneity and any such additional differentiation between hospitals only increases the NiN price for each hospital holding its utility to its patients fixed. GNT estimate the outside option share at 0.014 (table 2). In a symmetric Logit model, this implies $v = 1.86$. Using this, we calculate $S(10)$, $S(11)$ and K and from these p^N as a function of β and p_1 , p^D and p^R as a function of both β and δ .

Figure 5.1 plots the predicted equilibrium prices for all valid β values from the NiN and RSN models for selected δ . The NiN price “takes off” once $\beta > 0.6$ and the market breaks down (NiN price exceeds surplus) whenever $\beta > .85$. The effect of β is more contained in the RSN model as long as firms are at least somewhat patient ($\delta \geq 0.5$).

If the hospitals are at least as patient as usually assumed patience ($\delta = .8$), the model is not very sensitive to small changes in δ or β . The highest price in this case ($\beta = 1$) is 2.07, leaving a little over half of the total surplus to the insurer or patients. At the limit of perfectly patient firms, each hospital captures its value to the average patient in the market, $p^R \approx v = 1.86$ while the insurer captures all the gains from multiple hospitals ($\log(11) \approx 2.4$).

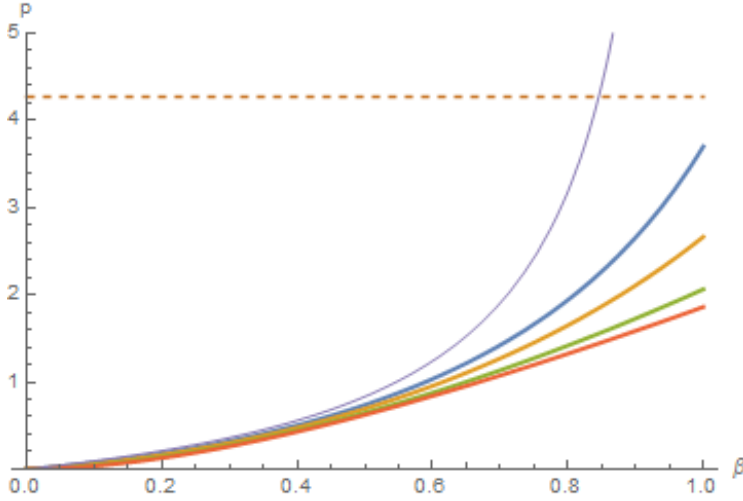


Figure 5.1: Price in the Logit example for $n = 11, v = 1.86$ as a function of the hospital bargaining power β . The dashed line is the full network value per patient, $\log(n \cdot e^v + 1)$. The higher, thin, line is the NiN price. The other, thick lines are the RSN solution with discount factors, for highest line to lowest, of 0.3, 0.5, 0.8 and 1.

5.4 Implication for Merger Evaluation

The workhorse models of hospital-insurer bargaining rarely predict a breakdown in negotiation. As a result, hospital merger evaluation tends to focus on the impact on negotiated prices between insurers and hospitals, assuming that the network structure remains unaffected. While prices have significant implications on the division of surplus, the main implication for total welfare stems from the reduction in health coverage due to the price increase.

In this section we follow this standard approach and consider the implication of using the RSN model for estimation relative to the NiN model when evaluating a merger between two hospitals A and B . As above, we assume that the RSN model estimation obtains $\delta \rightarrow 1$ so that the two models differ as much as possible.

The two-hospital example of section 2 shows that the NiN model would predict that prices will decrease after a merger whenever each hospital's price is higher than its value as an alternative. This finding would support approving the merger. In contrast, the RSN model for the same example must predict that prices increase, as there is only one hospital and the insurer cannot use the order of negotiations as a threat.

With more hospitals in the RSN model the punishment price for each hospital depends roughly on its average per-patient value as a single hospital. If the two merging hospitals are substitutes, the merger increases this value as the merged hospital provides each patient the max of the utility from being served either hospitals. The NiN model may predict a lower price effect and even a price decrease from a merger. Therefore the RSN model with high δ may suggest rejecting a merger while the NiN implies the opposite conclusion.

Summarizing, the NiN and RSN models would qualitatively differ in their prediction of post-merger prices whenever the hospitals considering merger are:

1. Substitutes: $q_{A,B}$ and $q_{B,A}$ are relatively high
2. Have pre-merger prices that are higher than the per-patient value the hospitals generate as a second alternative: $p_A > v_{B,A}$, $p_B > v_{A,B}$
3. The pre-merger price of alternative hospitals (j) is lower on average than p_B and p_A

The first two conditions imply that the NiN model would predict that prices are higher as a result of the bilateral negotiation. That is, if the hospitals were the only two in the network, the NiN model would predict the merger would decrease prices. This is because, in NiN, each hospital A and B as an alternative actually increases the pre-merger price of the other hospital. The third condition implies that the effect would decrease when considering a third alternative to two hospitals after they merge.

A standard but important caveat is that the estimation cannot, without additional assumptions, predict any changes the merger may generate on the underlying costs or bargaining power. In our setting, adopting different assumptions on this dimension may intensify the effect of any possible estimation bias in those parameters on the resulting conclusion. For example, if one assumes that β is unaffected by the merger, a merger between two hospitals with very low β (as estimated in specification 2 in GNT) cannot have significant negative price implications, and any cost savings from the merger will also be passed through to the patients. In contrast, if the estimated costs are much lower than they really are (as in specification 1 for patients of three insurers) it is difficult to accept any cost-saving arguments in favor of the merger, while those may well exist.

6 Conclusion

This paper has focused on the analysis of multilateral bargaining settings in which the total surplus from a suppliers’ network is non-linear in the (equilibrium) quantity sold by each supplier. This is particularly common when an intermediary (insurer or purchasing department) bargains with suppliers (hospitals or input producers) on behalf of downstream users.

We showed that if the intermediary cannot commit to a specific network, suppliers can charge unit prices that surpass the unit value because the intermediary must also consider the potential negative surplus from directing its users to their second-best suppliers.

This dynamic has significant normative implications. First and foremost, “forcing” the intermediary to commit not to reopen a failed negotiation guarantees that prices are below the users’ values. If this is not enforceable, encouraging profit (rather than surplus) maximization and reducing frictions in the downstream markets become very important.

Our proposed repeated sequential model describes the outcome that would result under a moderate (and reasonable) commitment power of the insurer—power which is sufficiently great to enforce sequential negotiations (and Nash reversion); the model can be used for estimation for both surplus and profit maximizing insurers.

Our analysis has several implications for empirical work. First, it identifies the important implications of the implicit “no-commitment” assumption common to the existing models. To the extent that this assumption is maintained, other assumptions – positive surplus, monotonic prices – and expectations – inefficient mergers increase prices – should be doubted. Second, it provides an alternative estimation approach based on the commitment assumption.

A Proofs

Note: Theorem 1 is proved in the text.

A.1 Preliminary Lemma

We first prove the following lemma. Suppose that the insurer is bargaining with hospital j and that:

1. The insurer's profit without j in the network is V_0 .
2. Adding j to the network at unit price p increases the insurer's profit by $K - p \cdot y$
3. The hospital's unit cost is c

Lemma 2. *If there is a price p that solves the bargaining game defined above, then it is:*

$$p = \beta \frac{K}{y} + (1 - \beta)c. \quad (\text{A.1})$$

The insurer's profit is $V_0 + (1 - \beta)(K - cy)$ and the hospital's profit per unit is $\beta \cdot \frac{K - cy}{y}$

Proof. Suppose the hospital and the insurer expect the hospital's quantity to be q . Then the bargaining problem is

$$\max_p [K - py]^{(1-\beta)} [x(p - c)]^\beta \quad (\text{A.2})$$

The interior solution must satisfy:

$$(1 - \beta) y [K - py]^{-\beta} [x(p - c)]^\beta = \beta \cdot x \cdot [K - py]^{(1-\beta)} [x(p - c)]^{(\beta-1)} \quad (\text{A.3})$$

Simplifying:

$$p \cdot [(1 - \beta) y + \beta y] = \beta K + c \cdot (1 - \beta) y \quad (\text{A.4})$$

Simplifying again obtains the desired price. the insurer and hospital profits are straightforward. \square

A.2 Two Hospitals Sequential Solution

Consider the negotiation with the second hospital given the first hospital's price p_1 . Denote by $q_{j,i}$ the number of patients that would go to hospital j if it was in the network but instead go to hospital $i \neq j$ because j isn't in the network.

Without hospital 2, $V^1 = F_1 - (q_1 + q_{2,1})p_1$. For any price p_2 the insurer's value is $V^{1,2} = F_{1,2} - q_1p_1 - q_2p_2$. Applying Lemma 2 for the bargaining gain $V^{1,2} - V^1$ obtains:

$$\begin{aligned} p_2 &= \beta \frac{F_{1,2} - F_1 + p_1q_{2,1}}{q_2} + (1 - \beta)c_2 \\ V^{1,2} &= F_1 - (q_1 + q_{2,1})p_1 + (1 - \beta)(F_{1,2} - F_1 + p_1q_{2,1} - q_2c_2) \\ &= (1 - \beta)(F_{1,2} - q_2c_2) + \beta F_1 - p_1(q_1 + \beta q_{2,1}). \end{aligned} \quad (\text{A.5})$$

Observe that as $q_{2,1} > 0$, p_2 is increasing in p_1 . Moreover, as by assumption it is surplus increasing to add the second hospital, $F_{1,2} - F_1 > c_2q_2$ and so $p_2 > c_2$.

Without hospital 1, $V^2 = (1 - \beta)(F_2 - (q_2 + q_{1,2})c_2)$. Therefore, the bargaining gain from 1 is

$$V^{1,2} - V^2 = \beta(F_1 - p_1q_{2,1}) + (1 - \beta)(F_{1,2} - F_2 + q_{1,2}c_2) - p_1q_1. \quad (\text{A.6})$$

Again applying lemma 2:

$$p_1 = \beta \frac{\beta F_1 + (1 - \beta)(F_{1,2} - F_2 + q_{1,2}c_2)}{q_1 + \beta q_{2,1}} + (1 - \beta)c_1. \quad (\text{A.7})$$

To see that $p_1 > c_1$, set $c_2 = 0$ (p_1 increases in c_2). Again, as long as $F_{1,2} - F_2 > q_1c_1$ (adding the hospital to the network is efficient) and $F_1 > c_1 \cdot (q_1 + q_{2,1})$ (the hospital is efficient in a network by itself), we have that $p_1 > c_1$.

We can now place p_1 in $V^{1,2}$ from equation A.5 to obtain the insurer's surplus:

$$\begin{aligned} V^{1,2} &= (1 - \beta)(F_{1,2} - c_2q_2) + \beta F_1 - (1 - \beta)[c_1(q_1 + \beta q_{2,1}) \\ &\quad - \beta(1 - \beta)(F_{1,2} - F_2 + q_{1,2}c_2) - \beta^2 F_1] \end{aligned} \quad (\text{A.8})$$

Simplifying:

$$\begin{aligned} V^{1,2} &= (1 - \beta)(F_{1,2} - c_2q_2 - c_1q_1) + \beta(1 - \beta)[F_1 - c_1q_{2,1} + F_2 - c_2q_{1,2} - F_{1,2}] \\ &= (1 - \beta)^2(F_{1,2} - c_2q_2 - c_1q_1) + \beta(1 - \beta)[F_1 - c_1(q_1 + q_{2,1}) + F_2 - c_2(q_2 + q_{1,2})] \end{aligned} \quad (\text{A.9})$$

A.3 J Hospitals Sequential Solution

Note that, with a slight abuse of notation, we use J to refer to the number of hospitals, the set of hospitals and the last hospital in the negotiation. We use the following notation for this section:

1. $\{1, \dots, J\}$ is the list of all relevant hospitals, in order of negotiation.
2. $V(\{p_1, \dots, p_i\}; \{i+1, \dots, J\})$ is the expected value on equilibrium for the insurer after negotiating with the first i hospitals and still having all the rest to negotiate with.
3. The value for a patient that would go to hospital j if it was in the network but instead goes to hospital $i \neq j$ because j isn't in the network is $v_{j,i}$.
4. The number of patients that would go to hospital j if it was in the network but instead go to hospital $i \neq j$ because j isn't in the network is $q_{j,i}$.

We make the following simplifying assumption: All consumers would leave the insurer if the two top hospitals in their list wouldn't be in the network. That is, there is no "third choice" hospital. Formally, if only hospital j leaves the network, the insurer loses $q_j - \sum_{k \neq j} q_{j,k}$ patients. If hospitals j and i leave the network, then the insurer loses $q_j - \sum_{k \neq j,i} q_{j,k} + q_i - \sum_{k \neq j,i} q_{i,k}$ patients.

Lemma 3. *The expected value for the insurer after negotiating with i of J hospitals and obtaining prices $\{p_1, \dots, p_i\}$ is given by:*

$$\begin{aligned}
 V(\{p_1, \dots, p_i\}; \{i+1, \dots, J\}) = & \sum_{j=1}^i \left[q_j(v_j - p_j) + \beta \sum_{k=i+1}^J q_{k,j}(v_{k,j} - p_j) \right] \\
 & + (1 - \beta) \sum_{j=i+1}^J \left[q_j(v_j - c_j) + \beta \sum_{k=i+1, k \neq j}^J q_{k,j}(v_{k,j} - c_j) \right].
 \end{aligned} \tag{A.10}$$

The price for insurer number i in the order of negotiations is given by:

$$\begin{aligned}
 p_i = & \beta \frac{q_i v_i + \beta \sum_{k=i+1}^J q_{k,i} v_{k,i} - \sum_{j=1}^{i-1} q_{i,j}(v_{i,j} - p_j) - (1 - \beta) \sum_{k=i+1}^J q_{i,k}(v_{i,k} - c_k)}{q_i + \beta \sum_{k=i+1}^J q_{k,i}} \\
 & + (1 - \beta) c_i
 \end{aligned} \tag{A.11}$$

Proof. By backward induction. Suppose the hospitals are ordered $1..J$. The value after J negotiations is: $V(\{p_1, \dots, p_J\}; \emptyset) = \sum_{j=1}^J q_j(v_j - p_j)$. When negotiating with the last hospital, the insurer's outside option is

$$V(\{p_1, \dots, p_{J-1}\}; \emptyset) = \sum_{j=1}^{J-1} [q_j(v_j - p_j) + q_{J,j} \cdot (v_{J,j} - p_j)] .$$

The gain from bargaining with J is therefore

$$\begin{aligned} W_J(\{p_1, \dots, p_{J-1}\}; \emptyset) &= V(\{p_1, \dots, p_J\}; \emptyset) - V(\{p_1, \dots, p_{J-1}\}; \emptyset) \\ &= q_J \cdot (v_J - p_J) - \sum_{j=1}^{J-1} q_{J,j} \cdot (v_{J,j} - p_j) . \end{aligned}$$

Applying Lemma 2:

$$p_J = \beta \left(v_J - \frac{\sum_{j=1}^{J-1} q_{J,j}(v_{J,j} - p_j)}{q_J} \right) + (1 - \beta)c_j .$$

Then we can place p_J to obtain the value after negotiating only with the $J - 1$ hospitals:

$$\begin{aligned} V(\{p_1, \dots, p_{J-1}\}; \{J\}) &= \sum_{j=1}^{J-1} q_j(v_j - p_j) + q_J v_J - q_J p_J \\ &= \sum_{j=1}^{J-1} [q_j(v_j - p_j) + \beta q_{J,j}(v_{J,j} - p_j)] + (1 - \beta)q_J(v_J - c_J) . \end{aligned}$$

We now have a formulation for both p_J and the insurer's value given the first $(J - 1)$ prices. Moving to the second to last bargaining game, if the bargaining succeeds and price is p_{j-1} the value to the insurer is given in $V(\{p_1, \dots, p_{J-1}\}; \{J\})$ above. If the bargaining fails, we can then continue to the J negotiation step with only the first $(J - 2)$ hospitals, accounting for the patients that would have went to hospital $(J - 1)$ but instead go to either an earlier hospital or to hospital J . The resulting

value to the insurer is

$$\begin{aligned} V(\{p_1, \dots, p_{J-2}\}; \{J\}) &= \sum_{j=1}^{J-2} [q_j(v_j - p_j) + q_{J-1,j} \cdot (v_{J-1,j} - p_j) + \beta q_{J,j}(v_{J,j} - p_j)] \\ &\quad + (1 - \beta) (q_J(v_J - c_J) + q_{J-1,J} \cdot (v_{J-1,J} - c_J)) . \end{aligned}$$

The gain from bargaining with $J - 1$ is therefore:

$$\begin{aligned} W_{J-1}(\{p_1, \dots, p_{J-1}\}; \{J\}) &= V(\{p_1, \dots, p_{J-1}\}; \{J\}) - V(\{p_1, \dots, p_{J-2}\}; \{J\}) \\ &= q_{J-1} \cdot (v_{J-1} - p_{J-1}) + \beta q_{J,J-1}(v_{J,J-1} - p_{J-1}) - (1 - \beta) q_{J-1,J} \cdot (v_{J-1,J} - c_J) \\ &\quad - \sum_{j=1}^{J-2} q_{J-1,j} \cdot (v_{J-1,j} - p_j) . \end{aligned}$$

Applying Lemma 2:

$$p_{J-1} = \beta \frac{q_{J-1}v_{J-1} + \beta q_{J,J-1}v_{J,J-1} - (1 - \beta)q_{J-1,J}(v_{J-1,J} - c_J) - \sum_{j=1}^{J-2} q_{J-1,j}(v_{J-1,j} - p_j)}{q_{J-1} + \beta q_{J,J-1}} + (1 - \beta)c_{J-1}$$

Placing p_{J-1} in $V(\{p_1, \dots, p_{J-1}\}; \{J\})$:

$$\begin{aligned} V(\{p_1, \dots, p_{J-2}\}; \{J-1, J\}) &= \sum_{j=1}^{J-2} \left[q_j(v_j - p_j) + \beta \sum_{k=J-1}^J q_{k,j}(v_{k,j} - p_j) \right] \\ &\quad + (1 - \beta) \sum_{j=J-1}^J \left[q_j(v_j - c_j) + \beta \sum_{k=J-1, k \neq j}^J q_{k,j}(v_{k,j} - c_j) \right] \end{aligned}$$

To generalize, suppose that for a “future hospitals” list of length $J - i$, the expected value for the insurer given prices $\{p_1, \dots, p_i\}$ is given by:

$$\begin{aligned} V(\{p_1, \dots, p_i\}; \{i+1, \dots, J\}) &= \sum_{j=1}^i \left[q_j(v_j - p_j) + \beta \sum_{k=i+1}^J q_{k,j}(v_{k,j} - p_j) \right] \\ &\quad + (1 - \beta) \sum_{j=i+1}^J \left[q_j(v_j - c_j) + \beta \sum_{k=i+1, k \neq j}^J q_{k,j}(v_{k,j} - c_j) \right] . \end{aligned}$$

Then:

$$\begin{aligned} V(\{p_1, \dots, p_{i-1}\}; \{i+1, \dots, J\}) &= \sum_{j=1}^{i-1} \left[q_j(v_j - p_j) + q_{i,j}(v_{i,j} - p_j) + \beta \sum_{k=i+1}^J q_{k,j}(v_{k,j} - p_j) \right] \\ &\quad + (1 - \beta) \sum_{j=i+1}^J \left[q_j(v_j - c_j) + q_{i,j}(v_{i,j} - c_j) + \beta \sum_{k=i+1, k \neq j}^J q_{k,j}(v_{k,j} - c_j) \right] . \end{aligned}$$

So the bargaining with i is over:

$$\begin{aligned} & V(\{p_1, \dots, p_i\}; \{i+1, \dots, J\}) - V(\{p_1, \dots, p_{i-1}\}; \{i+1, \dots, J\}) \\ &= q_i(v_i - p_i) + \beta \sum_{k=i+1}^J q_{k,i}(v_{k,i} - p_i) - \sum_{j=1}^{i-1} q_{i,j}(v_{i,j} - p_j) - (1 - \beta) \sum_{j=i+1}^J q_{i,j}(v_{i,j} - c_j). \end{aligned}$$

Again applying Lemma 2 obtains equation A.11.

We can now derive $V(\{p_1, \dots, p_{i-1}\}; \{i, \dots, J\})$ to complete the recursion. This can be done in two ways (that of course provide the same answer as above): either place p_i in $V(\{p_1, \dots, p_i\}; \{i+1, \dots, J\})$ or apply Lemma 2. Here are the steps using the first approach:

$$\begin{aligned} V(\{p_1, \dots, p_{i-1}\}; \{i, \dots, J\}) &= \sum_{j=1}^{i-1} \left[q_j(v_j - p_j) + \beta \sum_{k=i+1}^J q_{k,j}(v_{k,j} - p_j) \right] \\ &\quad + q_i v_i + \beta \sum_{k=i+1}^J q_{k,i} v_{k,i} - (q_i + \beta \sum_{i+1}^J q_{k,i}) p_i \\ &\quad + (1 - \beta) \sum_{j=i+1}^J \left[q_j(v_j - c_j) + \beta \sum_{k=i+1, k \neq j}^J q_{k,j}(v_{k,j} - c_j) \right] \end{aligned}$$

Placing p_i the second line changes to:

$$(1 - \beta) q_i(v_i - c_i) + (1 - \beta) \beta \sum_{k=i+1}^J q_{k,i}(v_{k,i} - c_i) + \beta \sum_{j=1}^{i-1} q_{i,j}(v_{i,j} - p_j) + \beta(1 - \beta) \sum_{j=i+1}^J q_{i,j}(v_{i,j} - c_j)$$

Moving the β term to the first line we get and the $(1 - \beta)$ term to the third line:

$$\begin{aligned} V(\{p_1, \dots, p_{i-1}\}; \{i, \dots, J\}) &= \sum_{j=1}^{i-1} \left[q_j(v_j - p_j) + \beta \sum_{k=i}^J q_{k,j}(v_{k,j} - p_j) \right] \\ &\quad + \beta(1 - \beta) \left[\sum_{k=i+1}^J q_{k,i}(v_{k,i} - c_i) + \sum_{j=i+1}^J q_{i,j}(v_{i,j} - c_j) \right] \\ &\quad + (1 - \beta) \sum_{j=i}^J \left[q_j(v_j - c_j) + \beta \sum_{k=i+1, k \neq j}^J q_{k,j}(v_{k,j} - c_j) \right] \end{aligned}$$

Collecting the $\beta(1 - \beta)$ terms in the second and third lines obtains the formulation in equation A.10. \square

For the two hospital case:

$$V(\emptyset, \{1, 2\}) = (1-\beta) [q_1(v_1 - c_1) + q_2(v_2 - c_2) + \beta (q_{2,1}(v_{2,1} - c_1) + q_{1,2}(v_{1,2} - c_2))] .$$

Setting $F_{1,2} = q_1v_1 + q_2v_2$ and $F_j = q_jv_j + q_{i,j}v_{i,j}$ this exactly equals the result in A.9.

A.4 Proposition 2

Assume the sequential model. For any $\beta \in (0, 1)$, the surplus generated by each insurer is positive.

Proof. We use the notation of section A.3. Observing equation A.10, the expected value for the insurer before any negotiations is:

$$V(\emptyset, \{1, \dots, J\}) = (1 - \beta) \sum_{j=1}^J \left[q_j(v_j - c_j) + \beta \sum_{k \neq j} q_{k,j}(v_{k,j} - c_j) \right] .$$

By assumption, $v_j > v_{k,j} > c_j$ for all j and so the sum is positive.

□

A.5 Proposition 3

Assume the sequential model with two hospitals or with J symmetric hospitals, if $\beta \rightarrow 1$, negotiating later in the order increases a hospital's price and profits.

Proof. We use the notation of section A.3. The order of negotiations only affects the price (and not quantity), and so it is necessary and sufficient to prove that the hospital's price is higher if it is later in the negotiations. We consider only the limit as $\beta \rightarrow 1$. In the two hospital setting, A 's price if it is first

$$p_A^1 = \frac{q_A v_A + q_{B,A} v_{B,A}}{q_A + q_{B,A}} .$$

If A is second, it's price is

$$p_A^2 = \frac{q_A v_A + q_{B,A} (p_B^1 - v_{B,A})}{q_A}$$

Then

$$p_A^2 - p_A^1 = \frac{q_A q_{B,A}(v_A - v_{B,A}) + q_{A,B}(q_A + q_{B,A})(p_B^1 - v_{A,B})}{q_A(q_A + q_{B,A})} \quad (\text{A.12})$$

As $v_A > v_{B,A}$ by construction and $p_B^1 > v_{A,B}$ when $\beta \rightarrow 1$ (see p_A^1 above), the difference must be positive.

For J hospitals, using equation A.11:

$$\lim_{\beta \rightarrow 1} p_j = \frac{q_j v_j + \sum_{k=j+1}^J q_{k,j} v_{k,j}}{q_j + \sum_{k=j+1}^J q_{k,j}} + \frac{\sum_{k=1}^{j-1} q_{k,j}(p_k - v_{j,k})}{q_j + \sum_{k=j+1}^J q_{k,j}}$$

The first fraction increases if j increases as $v_j > v_{k,j}$. For the second fraction to increase, a sufficient condition is that $\lim_{\beta \rightarrow 1} p_k \geq v_{j,k}$.

This is immediate by induction:

$$\lim_{\beta \rightarrow 1} p_1 = \frac{q_1 v_1 + \sum_{k=2}^J q_{k,1} v_{k,1}}{q_1 + \sum_{k=2}^J q_{k,1}} \in (v_{k,1}, v_1) .$$

For any $j > 1$, if for all $k < j$ it holds that $p_k > v_{j,k}$, then $p_j - v_{j,k}$ is even larger than it is for p_1 which completes the proof.²¹

□

A.6 Proposition 4

Assume the repeated sequential model. If hospitals have sufficiently large bargaining power, the punishment price converges to the hospital's per patient value if it was the only hospital in the network: For every $\varepsilon > 0$ there is a $\bar{\beta} < 1$ such that for all $\beta > \bar{\beta}$, $|\hat{p}_j - \frac{F_{jh}}{q_j^{jh}}| < \varepsilon$

Proof. We use the notation of section A.3. Using equation A.11:

$$\lim_{\beta \rightarrow 1} p_1 = \frac{q_1 v_1 + \sum_{k=2}^J q_{k,1} v_{k,1}}{q_1 + \sum_{k=2}^J q_{k,1}} = \frac{F_{jh}}{q_j^{jh}}$$

²¹In fact, for J symmetric hospitals, suppose that $q_{i,j} = q_j \cdot \alpha$ then using software we find that for every $j > 3$:

$$p_{j+1} - p_j = (p_j - p_{j-1}) \cdot \frac{1 + \alpha(J - (j + 1))}{1 + \alpha(J - (j - 2))} \quad (\text{A.13})$$

□

A.7 Proposition 5

In the equilibrium of the repeated sequential model, hospital prices are given by equation 4.3:

$$p_j = \delta \hat{p}_j + (1 - \delta) \frac{W_j}{q_j}.$$

Proof. Given J hospitals in the network, the insurer's problem is:

$$\begin{aligned} \min_{p_1, \dots, p_J} \quad & \sum_{j=1}^J q_j p_j \\ \text{s.t. } \forall j : \quad & p_j q_j - W_j(1 - \delta) \geq \delta \hat{p}_j q_j \end{aligned} \tag{A.14}$$

As W_j is a linear combination of all the prices, the problem is a standard constrained linear optimization problem with J unknowns and J constraints and thus all constraints bind. □

A.8 Proposition 6

In the repeated sequential model, if hospitals are sufficiently patient, the insurer's per-period surplus (or profit) is strictly positive. If hospitals have sufficiently large bargaining power, the insurer's profit converges to $F_h - \sum_j \left(F_{jh} \cdot \frac{q_j}{q_{jh}} \right)$.

Proof. By proposition 5, hospital prices p_j are given in 4.3. If $\delta \rightarrow 1$ then this converges to \hat{p}_j . Then proposition 4 implies that $p_j < v_j$, proving the first sentence. The first claim follows directly from proposition 2. The second claim follows from proposition 4. □

B Logit Example

This provides the technical details for the Logit example in section 5.3. For completeness, we add a taste shock variance σ . The consumption utility with n hospitals is:

$$U(n) = \sigma \log(n \cdot e^{\frac{v}{\sigma}} + e^{\frac{u_0}{\sigma}}) = \sigma \log((n + e^{\frac{u_0 - v}{\sigma}}) e^{\frac{v}{\sigma}}) = v + \sigma \log(n + e^{\frac{u_0 - v}{\sigma}})$$

The share of the outside option is $s_0(n) = \frac{e^{\frac{u_0}{\sigma}}}{e^{\frac{u_0}{\sigma}} + n \cdot e^{\frac{v}{\sigma}}}$.

The share of each hospital is $s(n) = \frac{e^{\frac{v}{\sigma}}}{e^{\frac{u_0}{\sigma}} + n \cdot e^{\frac{v}{\sigma}}}$.

The insurance surplus is $V(n) = v + \sigma \log(n + e^{\frac{u_0 - v}{\sigma}}) - s(n) \sum_{j=1}^n p_j$.

The added value from each hospital is

$$\begin{aligned} V(n) - V(n-1) &= \sigma \log \left(\frac{n + e^{\frac{u_0 - v}{\sigma}}}{n + e^{\frac{u_0 - v}{\sigma}} - 1} \right) - s(n) \sum_{j=1}^n p_j + s(n-1) \sum_{j=1}^{n-1} p_j \\ &= \sigma \log \left(\frac{n + e^{\frac{u_0 - v}{\sigma}}}{n + e^{\frac{u_0 - v}{\sigma}} - 1} \right) + (s(n-1) - s(n)) \sum_{j=1}^{n-1} p_j - s(n) p_n \end{aligned} \quad (\text{B.1})$$

The bargaining problem with each hospital solves

$$\max_{p_n} (V(n) - V(n-1))^{1-\beta} \cdot p_n^\beta$$

And the simultaneous NiN solution for all hospitals is given by

$$\begin{aligned} p_n &= \beta \cdot \left[\frac{\sigma}{s(n)} \log \left(\frac{n + e^{\frac{u_0 - v}{\sigma}}}{n + e^{\frac{u_0 - v}{\sigma}} - 1} \right) + \frac{s(n-1) - s(n)}{s(n)} \sum_{j=1}^{n-1} p_j \right] \\ &= \beta K(n, v, u_0, \sigma) + \beta s(n-1) \left(\sum_{j=1}^{n-1} \right) p_j \end{aligned} \quad (\text{B.2})$$

Where we set $K(n, v, u_0, \sigma) \equiv \frac{\sigma}{s(n)} \log \left(\frac{n + e^{\frac{u_0 - v}{\sigma}}}{n + e^{\frac{u_0 - v}{\sigma}} - 1} \right) > 0$ and use the relation:

$$\begin{aligned} \frac{s(n-1) - s(n)}{s(n)} &= \frac{s(n-1)}{s(n)} - 1 \\ &= \frac{e^{\frac{u_0}{\sigma}} + n e^{\frac{v}{\sigma}}}{e^{\frac{u_0}{\sigma}} + (n-1) e^{\frac{v}{\sigma}}} - 1 \\ &= \frac{e^{\frac{v}{\sigma}}}{e^{\frac{u_0}{\sigma}} + (n-1) e^{\frac{v}{\sigma}}} = s(n-1) \end{aligned} \quad (\text{B.3})$$

The unique solution is symmetric and so can be derived by summing up all the equations for the n hospitals. Letting $S(n) \equiv n \cdot s(n) = 1 - s_0(n)$ denote the total

share of all the hospitals in the market:

$$\begin{aligned} p^N &= \beta K(n, v, u_0, \sigma) + \beta S(n-1)p^N \\ &= \frac{\beta}{1 - \beta S(n-1)} K(n, v, u_0, \sigma) \end{aligned} \quad (\text{B.4})$$

For the RSN model, we start from equation 5.9, with p^R the equilibrium price, p^D deviation price and p_1 the punishment price.

For the numerical example, using GNT, the outside option share is 0.014 (table 2), u_0 is normalized to 0 and there are $n = 11$ hospitals. The outside option market share implies $v = 1.86$. Using this, we calculate the outside option share with one less hospital, $s_0(10) = .015$ and for $\sigma = 1$, $K(11, 1.86, 0, 1) = 9 \cdot \log(\frac{11+e^{-1.86}}{10+e^{-1.86}}) = .845$. Placing in p^N above obtains

$$p^N = \frac{\beta}{1 - .985\beta} .845$$

To determine p^R we need to find the punishment price p_1 . For this, set $v_{i,j} = v$ and $q_{i,j} = s(n-1)$ in equation A.11:

$$\begin{aligned} p_1 &= \beta v \frac{1 + (n-1)s(n-1)(2\beta-1)}{1 + \beta(n-1)s(n-1)} = \beta v \frac{1 + S \cdot (2\beta-1)}{1 + \beta \cdot S} \\ &\approx 3.72 \cdot \frac{\beta^2}{1.015 + \beta} \end{aligned} \quad (\text{B.5})$$

Placing p_1 in equation 5.9 completes the example.

References

- Abreu, D. (1986, June). Extremal equilibria of oligopolistic supergames. *Journal of Economic Theory* 39(1), 191–225.
- Abreu, D. and M. Manea (2012, January). Bargaining and efficiency in networks. *Journal of Economic Theory* 147(1), 43–70.
- Agency for Healthcare Research and Quality (2015, May). Percent of private-sector enrollees that are enrolled in self-insured plans at establishments that offer health

- insurance by firm size and selected characteristics (Table I.B.2.b.1), year 1996-2013. Medical Expenditure Panel Survey Insurance Component Tables. Generated using MEPSnet/IC. Technical report.
- Bennett, E. (1997, May). Multilateral Bargaining Problems. *Games and Economic Behavior* 19(2), 151–179.
- Binmore, K., A. Rubinstein, and A. Wolinsky (1986, July). The Nash Bargaining Solution in Economic Modelling. *The RAND Journal of Economics* 17(2), 176–188.
- Capps, C., D. Dranove, and M. Satterthwaite (2003). Competition and Market Power in Option Demand Markets. *The RAND Journal of Economics* 34(4), 737.
- Cleverley, W. O. and P. C. Nutt (1984, April). The effectiveness of group-purchasing organizations. *Health Services Research* 19(1), 65–81.
- Collard-Wexler, A., G. Gowrisankaran, and R. S. Lee (2014, October). Bargaining in Bilateral Oligopoly: An Alternating Offers Representation of the. SSRN Scholarly Paper ID 2518731, Social Science Research Network, Rochester, NY.
- Crawford, G. S. and A. Yurukoglu (2012, April). The Welfare Effects of Bundling in Multichannel Television Markets. *American Economic Review* 102(2), 643–685.
- Davidson, C. (1988, July). Multiunit Bargaining in Oligopolistic Industries. *Journal of Labor Economics* 6(3), 397–422.
- De Fontenay, C. and J. S. Gans (2013, March). Bilateral Bargaining with Externalities. SSRN Scholarly Paper ID 591688, Social Science Research Network, Rochester, NY.
- Gal-Or, E. (1997, March). Exclusionary Equilibria in Health-Care Markets. *Journal of Economics & Management Strategy* 6(1), 5–43.
- Gal-Or, E. (1999, September). Mergers and Exclusionary Practices in Health Care Markets. *Journal of Economics & Management Strategy* 8(3), 315–350.

- Gaynor, M. and R. J. Town (2011). Chapter Nine - Competition in Health Care Markets. In M. V. Pauly, T. G. McGuire, and P. P. Barros (Eds.), *Handbook of Health Economics*, Volume Volume 2, pp. 499–637. Elsevier.
- Gowrisankaran, G., A. Nevo, and R. Town (2015). Mergers When Prices Are Negotiated: Evidence from the Hospital Industry. *American Economic Review* 105(1), 172–203.
- Handel, B. R. (2013, December). Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts. *The American Economic Review* 103(7), 2643–2682.
- Harsanyi, J. C. and R. Selten (1972, January). A Generalized Nash Solution for Two-Person Bargaining Games with Incomplete Information. *Management Science* 18(5-part-2), 80–106.
- Ho, K. (2009, March). Insurer-Provider Networks in the Medical Care Market. *The American Economic Review* 99(1), 393–430.
- Ho, K. and R. S. Lee (2015). Insurer Competition in Health Care Market.
- Horn, H. and A. Wolinsky (1988, October). Bilateral Monopolies and Incentives for Merger. *The RAND Journal of Economics* 19(3), 408–419.
- Iozzi, A. and T. Valletti (2014, August). Vertical Bargaining and Countervailing Power. *American Economic Journal: Microeconomics* 6(3), 106–135.
- Kalai, E. (1977, September). Nonsymmetric Nash solutions and replications of 2-person bargaining. *International Journal of Game Theory* 6(3), 129–133.
- Klein, D. (2015, April). Interview with Guy Arie.
- Lee, R., M. Whinston, A. Yurukoglu, and G. Crawford (2013). The Welfare Effects of Vertical Integration in Multichannel Television Markets. 2013 Meeting Paper 81, Society for Economic Dynamics.
- Lewis, M. S. and K. E. Pflum (2015). Diagnosing Hospital System Bargaining Power in Managed Care Networks. *American Economic Journal: Economic Policy* 7(1), 243–74.

- Manea, M. (2011, August). Bargaining in Stationary Networks. *The American Economic Review* 101(5), 2042–2080.
- Noe, T. H. and J. Wang (2004, July). Fooling All of the People Some of the Time: A Theory of Endogenous Sequencing in Confidential Negotiations. *The Review of Economic Studies* 71(3), 855–881.
- Rey, P. and J. Tirole (2013). Cooperation vs. collusion: how essentiality shapes co-opetition.
- Scanlon, W. J. (2002). Group purchasing organizations: pilot study suggests large buying groups do not always offer hospitals lower prices. United States General Accounting Office Testimony, U.S. Senate Subcommittee on Antitrust, Competition, and Business and Consumer Rights.
- Schneller, E. S. (2000). The value of group purchasing in the health care supply chain. Mimeo., Tempe, Arizona.
- Smith, J. C. and C. Medalia (2014). Health Insurance Coverage in the United States: 2013. Technical report, US Census Bureau.
- Sorek, G. (2015, April). Health Insurance and Competition in Health Care Markets. Technical report, Auburn University.
- Stole, L. A. and J. Zwiebel (1996, July). Intra-firm Bargaining under Non-binding Contracts. *The Review of Economic Studies* 63(3), 375–410.
- Town, R. and G. Vistnes (2001, September). Hospital competition in HMO networks. *Journal of Health Economics* 20(5), 733–753.
- Wedig, G. J. (2013, January). The Value of Consumer Choice and the Decline in Hmo Enrollments. *Economic Inquiry* 51(1), 1066–1086.