Physician Practice and MCO Negotiation

The impact of time sensitive supply and demand

Daniel Ludwinski

November 25, 2017

Abstract

When health care providers and managed care organizations (MCOs) bargain, the main tool providers have is the threat to refuse to be in the MCO’s network. In fact, anecdotal evidence indicates that a major mechanism that practices employ to maximize profits in the face of differing insurer reimbursements, limited capacity and stochastic demand is to choose insurers discriminately. Providers do not accept patients from every MCO, however, providers do not exclusively accept the most profitable MCO. In this paper, I apply these institutional facts to a Nash cooperative bargaining framework to develop a bargaining model that explicitly models the provider’s disagreement point with the MCOs. In doing this, I am able to solve analytically for the interdependence of prices between MCOs and add to previous bargaining models by making the value of a MCO to a provider more explicit. This model shows the

1 Colgate University. Please send all questions, comments or corrections to DcLudwinski@colgate.edu
impact of MCO market structure on prices. By introducing provider capacity constraints, I am able to model two important provider-side considerations: the risk capacity will be unused, and the risk that a low paying patient will displace a higher paying patient. Neither of these two effects have been previously captured in the bargaining literature, which typically has featured marginal costs as the limiting factor for providers contracting with MCOs. I also show how predictions in my model match empirical observations and estimates from other work. I demonstrate a strong negative association between MCOs’ market power and negotiated prices, and show that the degree of market level price differences predicted by this model is similar to what has been observed. Finally, recent empirical work has found that price increases for Medicare are positively associated with private MCOs’ prices and that this impact is stronger in areas with more concentrated insurers, and areas in which Medicare patients represent a larger share of the market. My model analytically makes these predictions and can explain the underlying mechanisms.
1 Introduction

Many markets feature stochastic and time sensitive consumer demand along with supplier capacity that is static in the short term and non-storable – use it or lose it. Everyday examples include the market for live performances such as concerts or sporting events, restaurants and airline tickets. A common feature of these types of markets is the use of price as a market clearing mechanisms. The price is allowed to vary with contemporaneous demand. More popular events and restaurants have higher equilibrium prices. Airlines rapidly vary prices to avoid having unsold seats. Without price flexibility the result is typically excess demand (sell outs) or excess capacity (empty seats).

Though rarely applied to this context, the market for physician services also features stochastic and time sensitive consumer demand along with static, non-storable provider capacity. However, the market for physician services has both supply and demand side factors that do not allow a similar demand clearing mechanism. Prices for physician services are quite rigid. Medicare, the largest insurance provider in the United States, sets prices nationally. These prices are non-negotiable, and providers that participate in Medicare are forbidden from balance billing\(^2\). Similarly, Medicaid prices are generally set by states and are also a take it or leave it proposition. Reimbursement rates between physicians and private insurers are typically set once per year through a complex and opaque process of bilateral negotiations.

\(^2\) “Balance billing” is the practice of billing a patient the difference between the provider’s charge and the payment amount from a third-party payer, such as an MCO or Medicare.
Furthermore, there are three factors that make patient demand for medical services particularly unresponsive to price. First, for non-preventative care, there are often no good substitutes available, which means the underlying demand for physician services is generally inelastic with respect to price. The current best estimates of price elasticity for healthcare services are around 0.2 (Manning et al 1987, Newhouse et al. 1993, Zweifel and Manning 2000, Ringel et al 2002). Secondly, a substantial portion of the cost of care is covered by insurance, which means patients face neither the true cost of care or even the price that is transacted between their managed care organization and their healthcare provider. The result is that even if patient demand was more price-elastic, the price effect would be muted. Finally, even if a patient was particularly cost sensitive, prices are often unknown and not easily discoverable prior to service (Rosenthal, Lu and Cram 2013). Therefore, the mechanism through which the market for medical services clears must be more complex than menu prices directly influencing consumer demand.

There is a body of literature on provider market power and MCO-provider bargaining. Studies have shown a wide variation in prices across providers and MCOs (Baker, Bundorf, Royalty and Levin 2014, Ginsburg 2010, Cooper et al 2015). For example, Baker et al (2014) find that for internal medicine, the 10th and 90th percentiles for the Herfindahl-Hirschman Index, HHI, a common measure of market concentration, are respectively 666 and 3,154, and for urology they find 3,316 and 7,215. Research has shown that a factor in this price variation is market power, both for hospitals and physicians (Kleiner, White and Lyons 2015, Dunn and Shapiro 2014).

However, this literature currently does not include the above-mentioned features which are the mechanism through which a provider can leverage market power to receive higher prices.
Empirical work has examined how one payer’s price impacts the bargained price for another payer, for example changes in Medicare’s prices impacting private prices, see Frakt (2011) for a review of the evidence for hospital cost-shifting, and White (2013) for a more recent study. Most models used in empirical work assume bargaining outcomes that are independent across MCO-provider pairs (Grennan 2013, Lewis & Pflum 2015) and thus price does not explicitly depend on the market structure of the MCOs.

The goal of this paper is to add to the existing literature by examining the previously described features of the market for medical services – stochastic, time sensitive consumer demand and static, non-transferable supplier capacity in the face of rigid price structures and inelastic consumer demand. I will explicitly model how they impact the bargaining relationship between multiple managed care organizations (MCOs) and healthcare providers.

This paper proceeds as follows: In section 2, I give more background and motivation to justify and support the development of my approach. I show how I am building on the relevant literature, and contrast my approach with was has been done previously. I then develop, in section 3, a model of the physician’s decision to accept or reject a Managed Care Organization (MCO), given the expected price with and without that MCO. I discuss the MCOs desire to contract with the provider (the demand side), before combining the two into a dynamic bargaining model which incorporates the model of physician behavior. In section 4, I present basic predictions from the model. Finally, in section 5, I present some numerical examples and simulations and compare my results with previous research.
2 Background & Related Literature

An important assumption made in this paper is that in the short and medium-term physician and practice supply is relatively fixed. For practices, the intuition is that the main production inputs of space, equipment, and support staff cannot be easily varied day to day or week to week. For individual physicians, the idea is that their services are labor intensive. Physician labor responds to a price increase with competing income and substitution effects. While this assumption can be relaxed, the main formation of the model assumes that the effects cancel out and there is no aggregate supply response to price.

This assumption is not contradicted by the current literature. In an important early work looking at physician behavior McGuire and Pauly (1991) provide a theoretical model to test whether physicians have a target income or seek to maximize profits. They found that the strength of physicians’ income effect controls their behavior (Gruber, Kim, Mayzlin 1999, Yip 1998 Mitchell, Hadley, Gaskin 2000). More recently Kantarevic, Kralj and Weinkauf (2008) used reforms to the physician threshold system in Ontario, Canada to study this empirically. They find that, as expected, both the income effect and substitution effects are present with the expected signs. However, for different services, different effects dominate and there is no predominant aggregate supply effect.

The interplay between a practice and multiple payers, including Medicare, is an important mechanism in the model. A branch of the literature has sought to explain the response of private prices to changes in Medicare prices. Hospital administrators have advocated for “cost-shift theory”, that is, lower prices from one insurer will need to be made up somewhere to meet cost, and will then be shifted to other insurers. While economists have generally been skeptical of this theory, there is disagreement (Ginsburg 2003). In a 2011 review of the
literature, Frakt finds some evidence that cost shifting may occur, however the effects seem to be mild. In a more recent White (2013) finds the opposite effect – lower Medicare rates in hospitals resulted in lower private rates.

For physicians, Clemens and Gottlieb (2017) found consistent positive effects on private payer rates from increases in Medicare payments. These effects are larger both when Medicare makes up a larger share of the market and also when insurers have more relative market power. Ketcham, Nicholson, Unur and Lawrence (2014) similarly finds a positive relationship.

There is large existing literature covering MCO bargaining with providers for inclusion in a network. Town and Vistnes (2001) and Capps, Dranove and Satterthwaite (2003) use a logit demand model to construct a patient’s willingness-to-pay for inclusion of a provider based on observed provider and patient characteristics. These papers established the WTP concept as a measure of market power as well as the connection between that measure, profits, and prices. While originally focused on hospitals, these models have recently been applied to physicians as well (Carlson et al 2013, Kleiner, White and Lyons 2015). These papers, however, employ a standard bilateral Nash bargaining model, which does not explicitly include or model the interdependence of prices. The models show the impact of market concentration on the provider side, but cannot speak to the impact on prices stemming from different configurations of MCO market power. Across markets, there is wide variation in the concentration of insurers. According to a 2014 study by the Government Accountability Office, the three largest insurers in Wisconsin’s large group insurance market had a combined 39 percent of the total enrollees, while for most other states (37) the three largest insurers had more than 80 percent of the total commercial market.
More recent research has incorporated more sophisticated bargaining models. Ho and Lee (2013) study the price impact of insurer consolidation, focused on two competing forces. Increased insurer competition lowers premiums. Lower premiums reduce the surplus available to split between hospital and insurers, resulting in reduced prices. However, increased insurer competition gives hospitals more leverage to raise prices. They specify a general bargaining model in which price is determined by insurers’ premiums and payments to other hospitals, and hospitals’ costs and reimbursements from other payers. Lewis and Plum (2014) also develop a hospital, MCO bargaining model. Their innovation is to separately look at bargaining position (value of the hospital or network) and bargaining position (ability to obtain a higher share of the surplus).

I add to these bargaining models by making the value of a MCO to a provider more explicit. By introducing capacity constraints, I am able to model two important provider-side considerations: the risk capacity will be unused, and the risk that a low paying patient will displace a higher paying patient. Neither of these two effects have been previously captured in the bargaining literature, which typically has featured marginal costs as the limiting factor for providers contracting with MCOs. My paper will look at price differences arising from relative differences in MCO size stemming directly from the two effects. The model I put forward will not address any price differences that arise from efficiencies, bargaining ability, asymmetric information, or any pass-through price effects from the consumer-MCO price negotiations.
3 Model of Practice MCO Negotiation

Below I develop a model of practice-MCO bargaining. I explicitly specify the benefit of contracting for both the MCO and the practice, and show how for a given provider the negotiated prices are interdependent for each MCO.

In the first section, I introduce the providers problem by specifying the value to a practice of accepting patients of a particular type (taking prices as given). While this can be generalized to include any patient types that can be observable and discriminated, the focus here is on patients from different MCOs. Every MCO $k$ has a price ($p_k$) and a propensity ($\lambda_i$) – which can be thought of as the probability that a patient from MCO $k$ takes a given time slot, given the provider accepts patients from all MCOs. The probability in practice will depend on the contracting decisions for each of the other providers, which is a major mechanism in the model.

This expected value of including plan type $k$ depends on the prices of other accepted MCOs, their propensities, and propensity that a given time slot is unfilled ($\lambda_0$). This gives the value of the MCO to the provider.

Second, I characterize the value of the provider to the MCO by using the option-demand framework, developed by Capps et al (2003), to characterize an MCO’s willingness-to-pay for a patient to have the provider in the network as a function of patient’s expected utility. For use in bargaining, this is converted from utils to dollars and standardized to WTP per time slot to be comparable to the value of the MCO to the provider.

Third, I use the willingness-to-pay and the provider’s expected value in a Nash bargaining framework. The MCO and provider reach a deal to include the provider in the MCO network
if there is a price between the lowest price the provider would accept, the expected value of a
time slot without the provider, and the highest price the MCO would pay, which is the
willingness-to-pay. If they do reach an agreement they choose a price that splits the gains from
inclusion by a constant fraction. Unlike previous work, the explicit specification of the
provider’s value function allows me to solve the system of equations and derive a formula for
equilibrium prices that is determined simultaneously, depends on the both the provider and
MCO competitive landscape. This approach allows me to speak to the predicted relationship
of prices across MCOs.

Finally, I discuss the implications from and dynamics in this bargaining framework
demonstrating predictions from the model.

**Provider's Selection of MCOs**

In markets for restaurants, airline flights and concerts the market clears through direct price
increases or decreases, and generally prices are uniform across consumers. In the market for
health services, price changes happen through negotiations that generally occur once per year.
The main threat that providers have in these negotiations is the threat not to accept an MCO’s
patients. In the exposition below, I will concentrate on the agent being the physician practice,
but a similar framework could characterize other types of providers’ negotiations with MCOs.

For new patients especially, the availability of a convenient time slot not too far in the future
can be a major determinate of choosing a doctor. Anecdotal evidence indicates that physicians
take payer-mix into account when deciding whether to accept patients from a low paying
insurer. For new managed care contracts, the Practice Management Resource Group
encourages practices to evaluate “How the added patients will impact your payer-mix. Will
these patients increase or decrease your expected collections? Will they displace higher paying patients?”

Similarly, a popular book “Mastering Patient Flow” discourages closing practices fully to new patients due to the fact that it will decrease the practice’s ability to alter its payer mix. The alternative suggestion to alleviate capacity issues is to end participation with insurance companies that pay less.

In the model I develop here, the physician practice, indexed by j, faces K types of patients which it can either choose to accept or not accept – while this can be generalized to include any patient types that can be observable and discriminated, the focus for this exposition will be on patients from different MCOs.

Each slot is then filled with a patient of type k with a probability \( Prob_{k,j} \). Also, with positive probability, the time slot is not filled (denoted by \( Prob_{0,j} \)). This probability can be thought of as being market or provider specific, and in a manner detailed below, these probabilities will depend on the set of MCOs with which the provider has a contract. This way of characterizing the value of a time slot is applicable to arrangements where the provider is compensated through a fee-for-service system, and less relevant for physicians who are strictly salaried, or are compensated through capacitated arrangements, that is one in which the physician receives a set amount per patient year. I make the simplifying assumption that, conditional on contracting with an MCO, the practice cannot discriminate between patient types through

---

3 http://www.medicalpmrg.com/payor-mix-analysis.html (last accessed April 17, 2014)
4 Woodcock, Elizabeth W. Mastering Patient Flow (MGMA, 2009) 3rd edition
other means. Therefore, the practice’s problem is to evaluate the payouts from each patient type and choose which MCOs to accept.\(^5\)

The physician wants to choose the mix of MCOs (k) to maximize revenue (= the expected value of the time slot):

\[
\max_{K_j} EV_{K_j} = \max_{K_j} \sum_{k \in K_j} \text{Prob}_{k,j} p_k
\]

**Probability of type k:**

If \(\text{Prob}_{k,j}\) is exogenous to the choice of \(K_j\) (no capacity constraints), then all plans will be included. We do not observe this because, in practice, being able to accept and schedule a patient is conditional on having a time slot available. Therefore, a patient type with a low expected value (i.e. a low-paying MCO) can take the capacity away from a patient type with a higher expected value (a high-paying MCO). Furthermore, if there were no chance that a slot was not filled (excess capacity) then there would be no reason to accept any plan except for the highest paying. The tradeoff then is balancing the probability that no one takes the slot, with the probability that a patient with a lower paying plan prevents the provider from being able to render services to a patient with a higher paying plan.

This tradeoff can be formalized by denoting the unconditional probability of patient type k (the probability if all types are included) by \(\lambda_k\). Let \(\lambda_0\) be the probability that there are no

---

\(^5\) In the Appendices I include several variations and extensions of the model. I explicitly discuss excess capacity (Appendix A), I explore an alternative formation of the provider problem using a Poisson distribution of patients (Error! Reference source not found.), And finally, I show how the inclusion of variable cost (\(\delta\)) or exogenous physician work hours (Appendix C) do not significantly change the model.
patients in that time-period, given that all patients are accepted. I term this average excess capacity.

In the appendix, I discuss provider capacity and specify how providers choose capacity given expectations about patient demand, the expected marginal cost and expected payment for patient (not conditioned patient type). Adjusting this capacity is costly and fixed in the short and medium term. This leads to an optimal average excess capacity, or the propensity that a given time slot is unfilled ($\lambda_0$).

With these parameters defined, given a provider accepts the set of plans $K_j$, the probability of patient type $k$ is:

$$Prob_{k,j} = \frac{\lambda_k}{\lambda_0 + \sum_{k \in j} \lambda_k}$$

**Expected Value of a Time Slot**

Therefore, the expected value of a time slot can be expressed as follows:

$$EV_{K_j} = \sum_{k \in K_j} \lambda_k p_k \left[ \frac{\lambda_0 + \sum_{k \in K_j} \lambda_k}{\lambda_0 + \sum_{k \in K_j} \lambda_k} \right]$$

(1.0)

Maximizing this leads to the rule that patients of type $\delta$ should be included iff:

$$p_{\delta,j} \geq \left[ \sum_{k \in K_j/\delta} \lambda_k p_k \right] \left[ \frac{\lambda_0 + \sum_{k \in K_j/\delta} \lambda_k}{\lambda_0 + \sum_{k \in K_j/\delta} \lambda_k} \right] = EV_{K_j}$$

It is notable that the decision to include a particular type of patient does not depend on how many patients there are of that type (propensity $\lambda_\delta$). All that matters is the comparison between the expected value of the patient compared to the expected value of the set of currently
accepted patients. This expected value is influenced by share of slots likely to be unfilled, so the sizes of the other MCOs matter. While it is a minor distinction, bargaining power for a large MCO does not necessarily stem from the fact that the MCO is large, but stems from the fact that the other MCOs are not “large enough”. The ability to withhold quantity is a useless threat if the provider is already at capacity.

**Prediction 1:** A provider \((j)\) will want to contract with a MCO \((\delta)\) if the expected value of a time slot without the provider is lower than the price the MCO is offering.

With the rule under which a provider accepts an MCO established, we can examine some of the other dynamics predicted by the model.

**Addition of an MCO, \(\delta\):**

Using this formulation, the increase in the expected value of a time slot if provider \(i\) adds an insurer, given other accepted insurers \(K\) and prices, is:

\[
V_i(\delta|K_j, P) = \left[ \lambda_0 + \sum_{k \in K_j} \lambda_k p_k \right] / \left[ \lambda_0 + \lambda_\delta + \sum_{k \in K_j} \lambda_k \right] - \left[ \sum_{k \in K_j} \lambda_k p_k \right] / \left[ \lambda_0 + \sum_{k \in K_j} \lambda_k \right]
\]

\[
= \frac{\lambda_\delta}{\lambda_0 + \sum_{k \in K_j} \lambda_k} \left( \lambda_0 (p_\delta - 0) + \sum_{k \in K_j} (p_\delta - p_k) \lambda_k \right) / \left( \lambda_0 + \lambda_\delta + \sum_{k \in K_j} \lambda_k \right)
\]

(2.0)

This is the weighted price difference between \(\delta\) and the existing prices, normalized to a time slot, and multiplied by the percent increase in \(\lambda\) that \(\delta\) brings.
Demand Side: MCO’s Willingness-to-Pay for a Provider

To be able to discuss prices further, and to be able to examine bargaining dynamics, I first must specify the underlying demand system from the MCO. I do this by leveraging the option demand framework developed by Capps, Dranove and Satterthwaite (2003), through which an MCO has a willingness-to-pay to include the provider in the network.

In their model, a patient $i$ has ex post (that is, after the revelation of a health diagnosis requiring treatment) expected utility for the services from provider $j$ given by the following form:

$$ U_{ij} = \alpha R_j + H'_j X_i + \tau_1 T_{ij} + \tau_2 T_{ij} X_i + \tau_3 T_{ij} R_j - \gamma (X_i) P_j (Z_i) + \epsilon_{ij} $$

$$ = U(H_j, X_i, T_{ij}) - \gamma (X_i) P_j (Z_i) + \epsilon_{ij} $$

where $H_j$ are the provider characteristics, $X_i$ are the patient characteristic and $T_{ij}$ is the geographical location of the patient in relation to the provider. If the error term is logit, and we assume there are no meaningful out of pocket cost differentials between providers, then a patient’s utility of having access to a network $G$ of providers is:

$$ V^{IU}(G, Y_i, Z_i, T_{ij}) = E \max_{g \in G} [U(H_g, Y_i, Z_i, T_{ij}) + \epsilon_{ig}] = \ln \left( \sum_{g \in G} \exp U(H_g, Y_i, Z_i, T_{ig}) \right) $$

And the additional utility derived from the inclusion of provider $j$ is:

$$ \Delta V^I_{jU}(G, Y_i, Z_i, \lambda_i) = \ln \left( \frac{1}{1 - s_j (H_j, Y_i, Z_i, T_{ij})} \right) $$

This is the willingness to pay, in utils, for patient $i$ to have provider $j$ in network $G$. The willingness for the MCO to pay to have the provider in the system is calculated by summing this additional utility over all of patients in the MCO. In order to be used for my purposes, and compared to price, this WTP is then normalized as a WTP per visit, and converted to dollars.
The willingness-to-pay is the highest price an MCO would pay to have a provider in the network.

It is important to note that even if we assume that patient preferences do not differ systematically across MCOs – that is preferences only differ through the observed characteristics included in the utility function – the willingness-to-pay measures for a given provider can be different. Two main things drive this difference - the MCO’s network and the composition of patients.

Both $\Delta WTP$ and $\lambda_0$ (average excess capacity) reflect a provider’s desirability, but it is important to recognize how they are different in this model. The difference is that in this formation $\Delta WTP$ is normalized to a patient time slot, to correspond to price, and therefore does not depend on the size of the population. In contrasts $\lambda_0$ depends on the interplay between the number of patients, the number of other practices, and the size of the practice. If the number of patients increased, with no change in characteristics, $\Delta WTP$ normalized to a patient time slot would not change but $\lambda_0$ would decrease.

**Provider-MCO Bargaining**

I now apply a bargaining framework between providers (j) and MCOs (ℓ) to the above assumptions. In the standard Nash-bargaining framework, parties choose a price that splits the bargaining surplus, normalized to a per time-period amount, with constant parameter $\alpha \in (0,1)$. The typical assumption is that price solves the following:

$$p_{\ell j} = \arg\max_{p_{\ell j}} \left( WTP_{\ell j} - p_{\ell j} \right)^{1-\alpha} \left( p_{\ell j} - d_{j} \right)^{\alpha}$$
Where $d_j$ is the disagreement point for provider $j$ and $\alpha$ is the “price Nash bargaining parameter.” The outcome of the bargain depends non-trivially on the disagreement point and my contribution is to explicitly model this as previously described, $d_j = EV_{K_j/\ell}$. This makes the bargaining process between MCOs and providers explicitly interdependent. This is in contrast to other papers which assume independent bilateral bargaining (Lewis and Pflum 2015).

In the Nash solution, the MCO and the provider split the surplus, and this construction leads to the following set of price equations for each MCO ($\ell$) provider ($j$) pair:

$$p_{\ell j} = (1 - \alpha)WTP_{\ell j} + \alpha EV_{K_j/\ell}$$

$$= (1 - \alpha)\Delta WTP_{\ell j} + \alpha \left[ \sum_{k \in K_j/\ell} \lambda_k p_k \right] / \left[ \lambda_0 + \sum_{k \in K_j/\ell} \lambda_k \right]$$

The previously defined term, $EV_{K_j/\ell}$, means that prices are interdependent within a provider and thus must be determined simultaneously. Because the $\lambda'$s are taken as given, for each provider $j$ we have $L$ equations with $L$ unknowns, where $L$ is the total number of MCOs, and thus one can explicitly solve the equilibrium prices.

In the following sections, I show the equilibrium prices for some configurations of insurers, and discuss how these prices are impacted by the underlying parameters: the $\lambda$'s, each MCOs WTP, and administratively set prices.

**Monopolist**

If insurer $\delta$ is a monopolist then $EV_{K_j/\delta}$ is 0, and the equilibrium price equation is:
\[ p_{\delta} = (1 - \alpha)\Delta WTP_{\delta j} + \alpha EV_{K_{j/\delta}} = (1 - \alpha) \Delta WTP_{\delta j} \]

This is effectively the lowest price possible between insurer \( \delta \) and provider \( j \).

**Two Private MCOs and Medicare**

Consider the situation with two private insurers (indexed by 1 and 2), and Medicare (indexed by \( m \)). Because they are administratively set, in this model Medicare prices are taken as exogenous. This leads to the following equilibrium prices\(^6\):

\[
p_1^* = \left(1 - \alpha^2 \frac{\lambda_1}{\Lambda - \lambda_2} \frac{\lambda_2}{\Lambda - \lambda_1}\right)^{-1} \left[(1 - \alpha)WTP_1 + \alpha \frac{\lambda_m}{\Lambda - \lambda_1} p_m \right. \\
+ \alpha \frac{\lambda_2}{\Lambda - \lambda_1} \left. \left(1 - \alpha\right)WTP_2 + \alpha \frac{\lambda_m}{\Lambda - \lambda_2} p_m \right]
\]

(4.0)

Where for expositional simplicity, I have defined \( \Lambda \equiv \lambda_0 + \lambda_1 + \lambda_2 + \lambda_m \)

This equation characterizes the prices as a function of the underlying parameters. Price is related to the provider competitive landscape through the willingness to pay measure, and the insurer competitive landscape through the number of insurers and their relative sizes. As a note, the case without Medicare is equation 4.0 with \( \lambda_m = 0 \).

---

\(^6\) Details in Appendix E.
Equilibrium price is:

\[ p_1^* = \left( 1 - \alpha^2 \frac{\lambda_1}{\Lambda - \lambda_2} \frac{\lambda_2}{\Lambda - \lambda_1} \right)^{-1} \]

1. Market concentration premium (MCP)

\[ (1 - \alpha)WTP_1 \]

Own MCO’s willingness-to-pay

3. First order impact of Medicare price

\[ +\alpha \frac{\lambda_m}{\Lambda - \lambda_1} p_m \]

4. Other MCO price impact rate

\[ +\alpha \frac{\lambda_2}{\Lambda - \lambda_1} p_m \]

5. Other MCO’s willingness-to-pay

\[ ((1 - \alpha)WTP_2 \]

6. Second order Medicare price

\[ +\alpha \frac{\lambda_m}{\Lambda - \lambda_2} p_m \]

To explain this equilibrium price, I have separated it into six parts in the above table. The first part is the term \( \left( 1 - \alpha^2 \frac{\lambda_1}{\Lambda - \lambda_2} \frac{\lambda_2}{\Lambda - \lambda_1} \right)^{-1} \) which I call the market concentration premium (MCP).

It is always equal to or greater than 1. It captures a provider’s ability to extra a higher price by playing the MCOs off each other. If the prices were independently negotiated, then MCP would be 1. It is highest when the MCOs are the same size and for a given \( \lambda_0 \) and \( \lambda_m \).

The second term is the MCO’s willingness-to-pay, multiplied by the WTP bargaining coefficient \((1 - \alpha)\). WTP is the value that MCOs puts on having access to the provider and can change through changes in the underlying characteristics of MCOs population or network.

\footnote{In the more general case, with more than 2 insurers the MCP is \( 1/\text{det}(A) \), where \( A \) is the matrix defined in the technical appendix.}
Due to the MCP, an increase in MCO 1’s WTP for provider j increases the price between provider j and MCO by more than the bargaining parameter \((1 - \alpha)\). Intuitively, one can envision the following process leading to larger increase:

1. Looking at equation 3.0, when MCO 1’s WTP increases there is an immediate impact of an increase in \(p_1\) as they split the now larger surplus and the provider’s share is \((1 - \alpha)\).

2. However, this impacts the bargained prices between the provider and other MCOs. Having secured this higher price, the provider’s threat point (the expected value without the other MCOs) has increased. Therefore, the provider can now go to other MCOs and demand a higher price.

3. Once the provider has received the higher prices from the other providers they can return to MCO 1, and the process continues.

The third part is the first order impact of the Medicare price. The first term of this is the bargaining parameter \(\alpha\), and the second term, \(\frac{\lambda_m}{\Lambda - \lambda_1}\), is Medicare’s expected share of time slots if provider j did not accept MCO 1 patients. This is multiplied by the Medicare price, so this term is the contribution of Medicare to the expected value of the provider without MCO 1.

In a similar manner, the fourth term is the expected share of time slots for MCO 2 without MCO 1 multiplied by the bargaining parameter \(\alpha\), \(\alpha \frac{\lambda_2}{\Lambda - \lambda_1}\). This term captures the degree to which a price change for MCO 2 impacts the price for MCO 1. However, because the price for MCO 2 is not exogenous, the term is not just \(p_2\).
The fifth and sixth reflect the impact of MCO 2’s price on MCO 1’s price. The fifth term is the WTP bargaining coefficient \((1 - \alpha)\) multiplied by MCO 2’s WTP. The sixth term is the same as the third, however, it is multiplied by the price propagation factor, \(\alpha \frac{\lambda_2}{\Lambda - \lambda_1}\). This is the second order impact of Medicare, that is the impact on MCO 1’s price that happens through Medicare prices impacting MCO 2’s price.

4 Comparative Statics
These equilibrium prices lead to the following comparative statics and predictions. The equation that I present are only for the case with two private MCO and Medicare, but the predictions should hold more generally:

**Prediction 2:** The share of increase in MCO k’s demand (WTP) captured by provider j will be greater than the bargaining parameter \((1 - \alpha)\), and will depends on the market shares of all MCOs and the propensity for provider j to have an unfilled time slot \((\lambda_0)\).

\[
\frac{\partial p_1^*}{\partial \text{WTP}_1} = \left(1 - \alpha^2 \frac{\lambda_1}{\Lambda - \lambda_2} \frac{\lambda_2}{\Lambda - \lambda_1}\right)^{-1} (1 - \alpha) \geq (1 - \alpha)
\]

This prediction flows directly from equation (5) and the fact that this derivative is greater than the base bargaining parameter of \(1 - \alpha\). This is a result of modeling the interdependence of prices. An increase in WTP for MCO 1 will have a first-order increase on MCO 1’s price, however, this increase in MCO 1’s price will have a second-order impact on MCO 2’s price, and this chance in MCO 2’s price will have a third-order impact on MCO 1’s price, etc. The increase in price above \(1 - \alpha\) is result of that process being infinitely repeated, and is the equilibrium price. This leads to the next prediction from the model:
Prediction 3: There is a positive relationship between MCO i’s demand for provider j and other MCO’s contracted price with that provider.

\[
\frac{\partial p_1^*}{\partial WTP_{2}} = \left(1 - \alpha^2 \frac{\lambda_1}{\Lambda - \lambda_2} \alpha \frac{\lambda_2}{\Lambda - \lambda_1} \right) > 0
\]  \hspace{1cm} (6)

The second-order effect, described above, is shown in equation (6). An increase in demand by MCO 1 for provider j increases the equilibrium between provider j and other MCOs.

Prediction 4: There will be positive relationship between Medicare prices and private prices.

\[
\frac{\partial p_1^*}{\partial p_m} = \left(1 - \alpha^2 \frac{\lambda_1}{\Lambda - \lambda_2} \alpha \frac{\lambda_2}{\Lambda - \lambda_1} \right) \left[\alpha - \frac{\lambda_m}{\Lambda - \lambda_1} + \alpha - \frac{\lambda_m}{\Lambda - \lambda_2} \alpha \frac{\lambda_2}{\Lambda - \lambda_1} \right] > 0
\]  \hspace{1cm} (7)

The equilibrium bargained price depends strongly on the disagreement point, which is modeled as \( EV_{K_j/\ell} = \left[ \sum_{k \in K_j/\ell} \lambda_k p_k \right] / \left[ \lambda_0 + \sum_{k \in K_j/\ell} \lambda_k \right] \) (from equation (1)). The magnitude of the impact is the product of the market concentration premium, and the sum of what can be thought of as the first-order impact of the change in Medicare’s price \( (\lambda_m/(\Lambda - \lambda_1)) \) and the second-order impact on MCO 1 \((\alpha \frac{\lambda_2}{\Lambda - \lambda_1}) \) of the impact of the change in the Medicare price on MCO 2’s price \((\alpha \frac{\lambda_m}{\Lambda - \lambda_2}) \).

In short, the price provider j can command from MCO 1 has increased with the increase in Medicare’s reimbursement because the providers expected value without MCO 1 has increased. The increase in the Medicare price has increased the disagreement point, and therefore surplus that the MCO and the provider are bargaining has decreased.
**Prediction 5:** Even with identical underlying demand (WTP), larger insurers will pay a lower price.

\[
p_1^*/p_2^* = \left[ 1 + \alpha \left( \frac{\lambda_2}{\lambda_0 + \lambda_2} \right) \right] / \left[ 1 + \alpha \left( \frac{\lambda_1}{\lambda_0 + \lambda_1} \right) \right] \quad (8)
\]

\[
\frac{\partial}{\partial \lambda_j} \frac{\lambda_j}{\lambda_0 + \lambda_j} = \frac{\lambda_0}{(\lambda_0 + \lambda_j)^2} > 0 \quad (9)
\]

For predictions 5 and 6, I am ignoring Medicare and fixing WTP for MCO 1 equal to WTP for MCO 2. The size premium is the price discount, compared to other MCOs, that a larger MCO can achieve from its relative size. This ratio is less than 1 if \( \lambda_1 + \lambda_2 (1 + \alpha) < \lambda_2 + \lambda_1 (1 + \alpha) \Rightarrow \alpha \lambda_2 < \alpha \lambda_1 \), which means that if insurer 1 is larger then insurer 1 will pay less. The mechanism for this effect is the expected value to the provider without insurer 1 is smaller than the expected value without insurer 2.

**Prediction 6:** The differences in prices between large and small MCOs will be more pronounced among markets or providers with more excess capacity.

\[
\frac{\partial}{\partial \lambda_0} \left( \frac{\partial}{\partial \lambda_j} \frac{\lambda_j}{\lambda_0 + \lambda_j} \right) = \frac{\lambda_j - \lambda_0}{(\lambda_0 + \lambda_j)^3} > 0 \quad \text{if} \quad \lambda_j > \lambda_0 \quad (10)
\]

With an increase in \( \lambda_0 \), the size premium increases \( \left( \frac{\partial}{\partial \lambda_0} p_1^*/p_2^* > 0 \right) \). This happens because with a higher level of excess capacity the disagreement point for the provider, which is MCO's...
threat to not contract, is lower. While this is true for both MCOs, the effect is larger for the bigger MCO.

5 Examples
In this final section of the paper, I compute some expected prices, as a share of the difference between willingness-to-pay and cost, for several configurations of MCOs. I also show how prices and expected values change with the parameters.

Many of my predicted effects match empirical observations in the literature. I should a strong association between MCOs HHI and prices, and the magnitude is compatible with the Dunn and Shapiro estimates (2012).

I predict market level price differences that are similar to what Baker, Bundorf, Royalty, and Levin observe (2014). My model also matches the findings in Clemens and Gottleib (2017) that Medicare’s influence will be strongest in areas with concentrated insurers, and larger when Medicare makes up a larger share of the market.

Georgia vs Alabama
In order to illustrate the predicted differences in price as a function of market dynamics, I use data on the insurance markets for Alabama and Georgia. These numbers come from data on covered lives from the Medical Loss Ratio reports, so I’m simplifying by ignoring Medicare, Medicaid and self-insured plans. I also assume that the willingness-to-pay is identical across the insurer-providers pairs and the propensity that a given time slot is unfilled ($\lambda_0$) is 0.2.
What is shown in the table below is the market shares of the top five insurers for Alabama and George, and the corresponding implied prices (as a multiple of WTP).

<table>
<thead>
<tr>
<th></th>
<th>Alabama Market Share</th>
<th>Price</th>
<th>Georgia Market Share</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>74%</td>
<td>0.6950</td>
<td>33%</td>
<td>0.7940</td>
</tr>
<tr>
<td>2</td>
<td>10%</td>
<td>0.7778</td>
<td>28%</td>
<td>0.7984</td>
</tr>
<tr>
<td>3</td>
<td>8%</td>
<td>0.7792</td>
<td>15%</td>
<td>0.8101</td>
</tr>
<tr>
<td>4</td>
<td>5%</td>
<td>0.7819</td>
<td>14%</td>
<td>0.8104</td>
</tr>
<tr>
<td>5</td>
<td>3%</td>
<td>0.7849</td>
<td>10%</td>
<td>0.8115</td>
</tr>
<tr>
<td>Wtd Avg</td>
<td></td>
<td>0.7162</td>
<td></td>
<td>0.8018</td>
</tr>
</tbody>
</table>

The predicted prices for each MCO leads to the following three observations.

First, the price ratio of the 5\(^{th}\) largest to the largest is 1.02 in Georgia and 1.13 in Alabama. Second, the insurer with 10% market share in George has a 4.5% higher price than the insurer with the 10% market share in Alabama. This is a function of the dominant player being able to command a lower price, which results in a lower threat point for the rest of the insurers. Finally, the weighted average price is significantly (11%) lower in Alabama than in Georgia. The magnitude of this predicted difference is very much in line with the findings Baker, Bundorf, Royalty, and Levin (2014) who found the price difference in office visits between high HHI and low HHI regions to be between 8% and 16%.

In Georgia, the top insurer is Humana, and the fifth largest is Aetna. I can also estimate the impact of the merger had it been approved. I must note, however, that this analysis does not take into consideration any pass-through effects from their ability to raise prices on the plan consumers who purchase the plans, or any strategic responses on by providers.
<table>
<thead>
<tr>
<th>Market Share</th>
<th>Price</th>
<th>Market Share</th>
<th>Price</th>
<th>Price Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>33%</td>
<td>0.7940</td>
<td>2</td>
<td>43%</td>
</tr>
<tr>
<td>2</td>
<td>28%</td>
<td>0.7984</td>
<td>3</td>
<td>15%</td>
</tr>
<tr>
<td>3</td>
<td>15%</td>
<td>0.8101</td>
<td>4</td>
<td>14%</td>
</tr>
<tr>
<td>4</td>
<td>14%</td>
<td>0.8104</td>
<td>5</td>
<td>10%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wtd Avg 0.8018</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

By merging with Humana, Aetna could cut their reimbursement prices by 4.4% and Humana can cut theirs by 2.0%. Overall, prices drop by 1.5%, with the other insurers dropping reimbursements by more than 0.5%.

**Impact of an increase in WTP**

This model predicts how a change in how one MCO values a provider will change the price for both that MCO (prediction 3) and other MCOs (prediction 4). To illustrate with a numerical example, I set the bargaining parameter, $\alpha$, at 0.5, the Medicare propensity, $\lambda_m$, 0.25, both the MCOs propensities, $\lambda_1$ and $\lambda_2$, 0.3, and the propensity of a time slot to be unfilled, $\lambda_0$, 0.15. With these levels, below are the corresponding partial effects of an increase in WTP for MCO 1:

$$\frac{\partial p_1}{\partial \Delta WTP_{1j}} = 0.5240 \times d\Delta WTP_{1j}$$

$$\frac{\partial p_2}{\partial \Delta WTP_{1j}} = 0.1129 \times d\Delta WTP_{1j}$$

A model that used the same base Nash-bargaining parameters, but which ignored the interdependence of prices would predict an increase in price of $0.5 \times d\Delta WTP_{1j}$ for MCO 1 and 0 for MCO 2. My model predicts a slightly larger increase in prices for MCO 1, approximately 5% ($0.524/0.5-1$) higher than the static model. But my model also predicts that
there will be a considerable change in the prices for MCO 2. In fact, the provider is able to raise the price for MCO 2 by about 20% of the increase for MCO 1 (0.1129/.05240).

Impact of an increase in Medicare Reimbursements
The model also predicts a strong positive relationship between Medicare prices and private prices (prediction 4). Using the same parameters above, the impact of an increase in the Medicare price on MCO 1’s price is:

$$\frac{\partial p_1}{\partial p_m} = 0.2273 * d_p_m$$

While this is a significant effect, the increase is much less than 1, and much smaller than observed by Clemens and Goettlieb (2017). However, the positive predicted effect is incompatible with the theory of hospital cost-shifting (Frakt 2011).

Impact of MCO Size Differences
My model also can speak directly to the relationship between the relative size of the MCO and the relative prices each MCO will pay (prediction 5).

The MCO-provider negotiated prices have been modeled as a function of the characteristics and needs of the MCO’s customers, and the other providers already in the MCO network (substitutability). However, the resulting willingness-to-pay, once normalized to patient-visit, does not factor in the bargaining power of the MCO that stems for the relative importance of that MCO to the particular provider. To see how this plays out numerically I have calculated a couple of scenarios in which I have set average excess capacity ($\lambda_0$, in the first column), and the size parameters for MCO 1 and MCO 2 ($\lambda_1$ and $\lambda_2$ in columns 3 and 4 respectively). From those three parameters, I calculate the prices. I am setting WTP1=WTP2=0.
<table>
<thead>
<tr>
<th>$\lambda_0$</th>
<th>$\lambda_1$</th>
<th>$\lambda_2$</th>
<th>Price1</th>
<th>Price2</th>
<th>Size Premium</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>0.65</td>
<td>0.25</td>
<td>0.71</td>
<td>0.77</td>
<td>8.2%</td>
</tr>
<tr>
<td>0.05</td>
<td>0.65</td>
<td>0.25</td>
<td>0.88</td>
<td>0.91</td>
<td>3.4%</td>
</tr>
<tr>
<td>0.1</td>
<td>0.65</td>
<td>0.25</td>
<td>0.80</td>
<td>0.85</td>
<td>5.6%</td>
</tr>
<tr>
<td>0.1</td>
<td>0.75</td>
<td>0.15</td>
<td>0.75</td>
<td>0.83</td>
<td>10.9%</td>
</tr>
<tr>
<td>0.1</td>
<td>0.85</td>
<td>0.05</td>
<td>0.63</td>
<td>0.78</td>
<td>24.1%</td>
</tr>
<tr>
<td>0.1</td>
<td>0.55</td>
<td>0.35</td>
<td>0.83</td>
<td>0.85</td>
<td>2.5%</td>
</tr>
</tbody>
</table>

The exact size premium depends non-trivially on the values of the parameters. However, the size premium is consistent and for large differences in size, considerable.

**Impact of Average Excess Capacity ($\lambda_0$)**

It is important to recognize how willingness-to-pay (WTP) and $\lambda_0$ (average excess capacity) differ in this model, as both reflect aspects of a provider’s desirability. In my model, WTP is normalized to a patient time slot, to correspond to price, and therefore it does not depend on the size of the population. Instead, it depends on patient and provider characteristics such as location, health status, etc). In contrast, average excess capacity ($\lambda_0$) depends on the interplay between the overall number of patients, the number of other practices, the propensity for a patient to choose the practice and the size of the practice. An increase in the total number of patients, without a corresponding increase in physicians, will not increase the patient normalized willingness-to-pay, but it will decrease $\lambda_0$ (average excess capacity). An increase in a practice’s capacity again, would not increase the patient normalized willingness-to-pay, but this will increase $\lambda_0$.

Below I provide a numeric example of how a change in $\lambda_0$ results in higher prices, even while ignoring any impacts from the increase in willingness-to-pay. Using the model of prices with
two insurers (no Medicare), the following table contains the corresponding equilibrium prices for two different configurations of market share, and two different values for $\lambda_0$, 0.2 (meaning that the underlying probability that a slot will be taken is 80%) and 0.1. $WTP_1$ and $WTP_2$ both are fixed at 1:

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>Price</th>
<th>$\lambda_0 = 0.2$</th>
<th>Price</th>
<th>$%$ Price Increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.40</td>
<td>0.7500</td>
<td>0.846</td>
<td>12.8%</td>
</tr>
<tr>
<td>2</td>
<td>0.40</td>
<td>0.7500</td>
<td>0.846</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>Price</th>
<th>$\lambda_0 = 0.1$</th>
<th>Price</th>
<th>$%$ Price Increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.64</td>
<td>0.6676</td>
<td>0.7693</td>
<td>15.2%</td>
</tr>
<tr>
<td>2</td>
<td>0.16</td>
<td>0.7543</td>
<td>0.8377</td>
<td>11.1%</td>
</tr>
</tbody>
</table>

Wtd Avg 0.6849 0.7830 14.3%

In both cases, with the MCOs have equal market share and where one is larger, there is a significant increase in price from the decrease in $\lambda_0$. The increase is smaller in the equal shares case, a 12.8% increase. With different shares, the larger MCO is forced to increase their reimbursement more than the smaller – 15.2% vs 11.1%. The weighted average price increase is 14.3%. These price increases do not stem from a higher willingness-to-pay for a timeslot, but are the result of providers being able to be firmer in their negotiations as there is a smaller probability that they will not find patients, given they are not accepting patients from an MCO.

Finally, the following charts provide a visualization of the price dynamics with two insurers. I show the price for each MCO (the lines) and the expected value of a time slot (the areas). I show these three values along one of three dimensions: the relative size of the insurers,
average excess capacity ($\lambda_0$) and the ratio of the insurers’ willingness-to-pay. Lastly, I compare Herfindahl-Hirschman Index (HHI) and price resulting from my model.

Varying the high-paying share

In Figure 1, the two insurers have a different WTP, meaning the provider is more valuable to one MCO than to the other. This could stem from the provider’s skill set matching up better with the needs of one MCO’s population, or it could result from convenience and distance. The WTP is set to 1 for the insurer that values the provider less and the WTP is 2 for the other insurer. The average excess capacity, $\lambda_0$, is held constant at 10%. What varies on the x-axis is the share of patients that are in the high paying MCO. The main insight from this graphic is that at the two extremes the provider accepts both patients, but in the middle the provider only accepts the higher-paying patient type. The intuition is that if the high paying share is “high enough” (in this example 25%) than the risk of a low paying patient crowding out a low paying patient is not worth the risk of having an empty slot. On the other end, as the MCO that is
willing to pay more has a higher share, it is able to use that market power to drive down their price. Eventually, the price is low enough that the cost of a low-paying patient crowding out a high-paying one is small.

**Varying Average Unused Slot ($\lambda_0$)**

In Figure 2, both the size of the MCO patient population and the MCO’s WTP are held constant at 2 and 1. What varies is $\lambda_0$. For low value of $\lambda_0$, the provider should only accept patients from the high-paying MCO, which is the grey portion of Figure 2.

The intuition is straightforward. As $\lambda_0$ falls, there is a smaller probability that there will be unused capacity if the provider drops the low paying MCO.
Varying the ratio of the MCOs’ willingness-to-pay

In the above graphic, the WTP ratios are varied in such a way as to keep the average WTP constant at 1. The patient populations of both MCOs are held constant and equal. The expected value to the provider decreases as the WTP ratio heads to one. The provider should accept patients from both MCOs unless the WTP ratio is higher than 1.82.

Herfandahl-Herman Index and Average Price
The Herfandahl-Herman Index (HHI) is a standard measure of industry concentration that is often used in industrial organization literature and used by government agencies tasked with enforcing antitrust law. The assumption is that more concentrated markets (higher HHI) on the producer side (in this case the providers) will result in higher prices and more concentrated markets on the purchaser side (in the case, the MCOs) will result in lower prices. By using my model, I can calculate simulate different arrangements of market structure and calculate the corresponding average negotiated price (as a share of willingness-to-pay). I do this for a
set off 2,500 various market shares with four MCOs, holding $\lambda_0$ constant. As shown in the graph below, there is a striking relationship between HHI and average price resulting from above model. The fitted quadratic equation is given by:

$$\text{price} = 0.8134 + 0.015 \; \text{HHI} - 0.3101 \; \text{HHI}^2$$

where HHI has been divided by 10,000 in the equation above for readability. In the figure below the fitted line is in black and the generated HHI, average price pairs are plotted in blue.

6 Conclusion

In this paper, I propose a structural bargaining model that is most readily applicable to MCO-healthcare provider bargaining. The main innovation of this model is to model the disagreement point explicitly for a provider not reaching an agreement with an MCO. In the model, the disagreement point is a function of the negotiated prices between other MCOs and the provider, as well as overall demand-side factors which play into willingness-to-pay and average excess capacity. In this way, the prices for each MCO are explicitly interdependent.
within providers. I am able to model this disagreement point by exploiting the fact that two large factors in provider-MCO bargaining are providers have a limited ability to service patients, and patient demand is time sensitive and variable.

Using this model, I demonstrate the conditions under which a provider will want to contract with a MCO and I analytically solves for how relative provider size and provider concentration impact the negotiated prices, and how price-interdependence leads to cross-price effects within a provider between MCOs. The magnitude and direction of the model’s predicted effects are validated by comparing predictions of model to previously observed statistics or estimated relationships, such as the average price difference between regions, the positive impact of an increase in Medicare prices on private MCO prices (including when that impact will be strongest). My model matches some previous findings, while providing a potential explanation for underlying causal mechanisms.

While this model is limited by the fact that I do not explicitly model concentration on the provider side, and do not take into consideration the impact of MCO concentration prices on premiums, it adds to our understanding of MCO bargaining, as the mechanisms of limited capacity and time-sensitive demand have not previously been incorporated into a structural MCO-provider bargaining framework. While this work is most easily applied to MCO-provider negotiation, this model could potentially be adapted to apply more closely to other industries as many markets face time sensitive demand and non-storable supply.
REFERENCES


Woodcock, Elizabeth W. Mastering Patient Flow (MGMA, 2009) 3rd edition


Appendix A: Capacity

In this appendix, I derive the average excess capacity ($\lambda_0$). Excess capacity is a profit maximizing strategy for firms when there are fixed cost associate with building that capacity and uncertainty about how many consumers will arrive in a given period of time.

First, let the firm (physician) have a belief about the expected value of a given unit of capacity ($EV_c$) which is expected net price (expected price minus expected variable cost). Second, denote capacity by $S$ (size) and let the cost for every unit of capacity be fixed at $c_S$. Finally, let the number of customers (patients) in a given time-period be approximated by a Poisson distribution with mean and variance $x$.

The firm then chooses capacity $S$ to maximize the following profit function:

$$\Pi = -c_S S + EV_p \sum_{i=0}^{S} \frac{e^{x_i}}{i!} + EV_p \left( \sum_{i=S+1}^{\infty} \frac{e^{x_i}}{i!} \right)$$

And the change in expected profit from an extra unit of capacity is:

$$\frac{\Delta \Pi}{\Delta C} = -c_S + EV_p \left[ S \left( \frac{e^{x_S}}{S!} - \frac{e^{x_{S+1}}}{(S + 1)!} \right) + \sum_{i=S+2}^{\infty} \frac{e^{x_i}}{i!} \right]$$

Therefore, the rule to maximize profit is to add a unit of capacity if (subject to positive overall profit):

$$c_S < EV_p \left[ S \left( \frac{e^{x_S}}{S!} - \frac{e^{x_{S+1}}}{(S + 1)!} \right) + \sum_{i=S+2}^{\infty} \frac{e^{x_i}}{i!} \right]$$
This provides the optimal level of capacity as a non-linear function of the unconditional average number of patients in a time-period (x) and the ratio between the cost of an extra unit of capacity and the expected value of a patient and also allows the average excess capacity ($\lambda_0$) to be calculated.

$$S^* = S\left(x, \frac{c_s}{EV_p}\right)$$

$$\lambda_0\left(x, \frac{c_s}{EV_p}\right) = \frac{1}{S^*} \sum_{i=0}^{S^*} (S^* - i) \frac{e^x x^i}{i!}$$

Optimal capacity is increasing in the unconditional mean (x) and decreasing in the cost/expected value ratio (holding constant EV higher cost of capacity will lead to less capacity). Lambda_0 is decreasing both in the unconditional mean and capacity, and increasing in the capacity cost to expected value ratio. To give an example, with a fixed cost-to-expected value ratio of 10%, a provider facing a patient Poisson distribution with a mean of 20 will have a capacity of 30 and an average excess capacity of 33.4% percent, while a provider facing a patient demand distribution with a mean of 455 will have a capacity of 500 and an average excess capacity of only 9.4% percent.

An important note on the definitions of a time-periods and capacity. For a restaurant, capacity can be thought of tables, for hospitals beds and for physicians, appointment slots. The time-period is the relevant period for which a consumer’s demand remains active. For a patient, the time-period should be thought of as a period in which once a patient realizes their health state, they can be flexible. Therefore, it will differ by type of service, and type of patient. The
relevant time-period for a cardiac intensive care unit may have a time-period of 30 minutes, while the correct time-period for primary care office may be several weeks.
Appendix B: Provider Problem Including Variable Costs

If variable costs are included then the expected value of a time slot is:

$$EV_{K,j} = \frac{\sum_{k \in K_j} \lambda_k (p_k - c_j)}{\lambda_0 + \sum_{k \in K_j} \lambda_k} = \frac{\sum_{k \in K_j} \lambda_k p_k}{\lambda_0 + \sum_{k \in K_j} \lambda_k} - \frac{\sum_{k \in K_j} \lambda_k c_j}{\lambda_0 + \sum_{k \in K_j} \lambda_k}$$

And the change in expected value of time slot from including patients of type $\delta$ is:

$$\left( \frac{\lambda_\delta (p_k - c_j) + \sum_{k \in K_j} \lambda_k (p_k - c_j)}{\lambda_\delta + \lambda_0 + \sum_{k \in K_j} \lambda_k} \right) - \left( \frac{\sum_{k \in K_j} \lambda_k (p_k - c_j)}{\lambda_0 + \sum_{k \in K_j} \lambda_k} \right)$$

This equation is very similar to equation used in the paper. Furthermore, the inclusion rule is very similar, it only now includes costs explicitly. This changes the amount of total surplus that the MCO and provider negotiate over, and therefore can change the predictions about the price. The underlying fact that provider’s threat point is the expected value without the MCO, however, remains mostly unchanged.
Appendix C: Time & Leisure in the Physician’s Problem

In what follows, I allow the hours worked for the physician to depend on the expected return to working.

\[ u_j(c, q) = \log \left( \sum_{k=1}^{K} p_k q_k \right) - \alpha_j \log \left( X - \sum_{k=1}^{K} q_k \right) \]

\[ q = \sum_{k=1}^{K} q_k \]

\[ \max_{K_j,q} E[u_j(q, K_j)] = \max_{K_j,q} \left( -\alpha_j \log(X - q) + \sum_{k \in K_j} q_k \cdot p_k \cdot Prob_k \right) \]

Note that price should be thought of not as the list of transacted price for that patient, but full net expected payment taking into considerations the cost of working with that type of patient or insurance company.

For simplicity let \( q_k = 1, \forall k \)

**FOC q:**

\[ \frac{\partial EU}{\partial q} = \frac{\alpha_j}{X - q} + \sum_{k \in K_j} p_k \cdot Prob_k = 0 \]

\[ q^* = X - \frac{\alpha_j}{\sum_{k \in K_j} p_k \cdot Prob_k} \]

Then using this to calculate the expected utility of accepting the set of patients \( K_j \):

\[ E[u_j(q^* | K_j)] = \alpha_j \log \left( X - \left( X - \frac{\alpha_j}{\sum_{k \in K_j} p_k \cdot Prob_k} \right) \right) + \left( X - \frac{\alpha_j}{\sum_{k \in K_j} p_k \cdot Prob_k} \right) \sum_{k \in K_j} p_k \cdot Prob_k \]
\[
= \alpha_j \log \left( \frac{\alpha_j}{\sum_{k \in K_j} p_k \cdot \text{Prob}_k} \right) + X \sum_{k \in K_j} p_k \cdot \text{Prob}_k - \alpha_j
\]

\[
= X \sum_{k \in K_j} p_k \cdot \text{Prob}_k - \alpha_j \left[ 1 - \log(\alpha_j) + \log \left( \sum_{k \in K_j} p_k \cdot \text{Prob}_k \right) \right]
\]

**Partial Derivatives for Physician Problem**

For the following derivatives prices are held constant.

**Hours Worked, wrt \( p_l \):**

\[
q^* = X - \alpha_j \frac{\lambda_0 + \sum_{k \in K_j} \lambda_k}{\sum_{k \in K_j} \lambda_k p_k} = X - \alpha_j \frac{\lambda_0 + \sum_{k \in K_j} \lambda_k}{\lambda_l p_l + \sum_{k \in K_{j/l}} \lambda_k p_k}
\]

\[
\frac{\partial q^*}{\partial p_l} = -\alpha_j \frac{\lambda_0 + \sum_{k \in K_j} \lambda_k}{\left( \sum_{k \in K_j} \lambda_k p_k \right)^2} = -\frac{\alpha_j}{\sum_{k \in K_j} \lambda_k p_k} \left( \frac{1}{\text{EV} (K_j)} \right) < 0
\]

**Hours Worked, wrt \( \lambda_l \):**

Note that an increase in \( \lambda_l \) can be interpreted as an increase in demand by patients of type \( l \).

\[
q^* = X - \alpha_j \frac{\lambda_0 + \sum_{k \in K_j} \lambda_k}{\sum_{k \in K_j} \lambda_k p_k} = X - \alpha_j \frac{\lambda_l}{\lambda_l p_l + \sum_{k \in K_{j/l}} \lambda_k p_k} - \alpha_j \frac{\lambda_0 + \sum_{k \in K_{j/l}} \lambda_k}{\lambda_l p_l + \sum_{k \in K_{j/l}} \lambda_k p_k}
\]

\[
\frac{\partial q^*}{\partial \lambda_l} = -\alpha_j \left[ \frac{\sum_{k \in K_{j/l}} \lambda_k p_k}{\left( \sum_{k \in K_j} \lambda_k p_k \right)^2} - \frac{p_l \left( \sum_{k \in K_{j/l}} \lambda_k \right)}{\left( \sum_{k \in K_j} \lambda_k p_k \right)^2} \right]
\]

\[
= -\alpha_j \left[ \frac{\sum_{k \in K_j} \lambda_k p_k}{\left( \sum_{k \in K_j} \lambda_k p_k \right)^2} - \frac{\sum_{k \in K_j} \lambda_k p_l}{\left( \sum_{k \in K_j} \lambda_k p_k \right)^2} \right]
\]
\[
= -\frac{\alpha_j}{\sum_{k \in K_j} \lambda_k p_k} \left[ 1 - p_l \frac{\sum_{k \in K_j} \lambda_k}{\sum_{k \in K_j} \lambda_k p_k} \right] \\
= -\frac{\alpha_j}{\sum_{k \in K_j} \lambda_k p_k} \left[ 1 - p_l \frac{1}{\mathbb{E}V(K_j)} \right] \\
= \frac{\alpha_j}{\sum_{k \in K_j} \lambda_k p_k} \left[ p_l - 1 \right]
\]

So, if \( p_l \) is higher than the expected value of the set then hours worked increases. Else, it decreases.

Since all included \( p_l \)'s must be higher than the expected value (assuming ability to discriminate on types), then for all \( l \), work increases with the number of patients.

\[
\frac{\partial q^*}{\partial \lambda_l} > 0
\]

If we're talking about \( \lambda_0 \), than \( p \) is 0 so because \( a_j \) is greater than 0 the derivative is negative (less work). So in this simple model, doctors work more in respect to positive demand shocks, and less in response to negative demand shocks (as expected). The substitution effect dominates the income effect.

**Patients Seen (wrt \( \lambda_l \)):**

Patients seen =

\[
q^* \left( 1 - \text{prob}_0 \right) = q^* \left( 1 - \frac{\lambda_0}{\lambda_0 + \sum_{k \in K} \lambda_k} \right) = q^* \left( \frac{\sum_{k \in K} \lambda_k}{\lambda_0 + \sum_{k \in K} \lambda_k} \right)
\]

\( q^* \) rises (number of slots), and patients per slot (fill rate) rises drops as well, so patients seen rises.
Expected Value (wrt $\lambda_l$):

$$\frac{\partial EV_{K_j}}{\partial \lambda_l} = p_l \left[ \frac{\lambda_0 + \sum_{k \in K_j/l} \lambda_k}{\lambda_0 + \sum_{k \in K_j} \lambda_k} \right] - \left[ \frac{\lambda_0 + \sum_{k \in K_j/l} \lambda_k p_k}{\lambda_0 + \sum_{k \in K_j} \lambda_k} \right]$$

$$= \frac{\lambda_0 p_l + \sum_{k \in K_j/l} \lambda_k p_l - \lambda_0 - \sum_{k \in K_j/l} \lambda_k p_k}{\left[ \lambda_0 + \sum_{k \in K_j} \lambda_k \right]^2}$$

$$= \frac{\sum_{k \in K_j} \lambda_k (p_l - p_k)}{\left[ \lambda_0 + \sum_{k \in K_j} \lambda_k \right]^2} + \frac{\lambda_0 (p_l - 1)}{\left[ \lambda_0 + \sum_{k \in K_j} \lambda_k \right]^2}$$

$$= \frac{1}{\sum_{k \in K_j} \lambda_k} \left[ \left( \frac{\lambda_0 + \sum_{k \in K_j} \lambda_k}{\lambda_0 + \sum_{k \in K_j} \lambda_k} \right) p_l - \left( \frac{\lambda_0 + \sum_{k \in K_j} \lambda_k}{\lambda_0 + \sum_{k \in K_j} \lambda_k} \right) p_k \right]$$

$$= \frac{1}{\sum_{k \in K_j} \lambda_k} \left[ p_l \frac{\lambda_0 + \sum_{k \in K_j} \lambda_k}{\lambda_0 + \sum_{k \in K_j} \lambda_k} - \sum_{k \in K_j} \lambda_k p_k \right]$$

$$= \frac{p_l - EV_{K_j}}{\sum_{k \in K_j} \lambda_k} > 0$$

Not surprisingly, an increase in demand increases the expected value of a time slot. The magnitude of the increase depends on the difference between the price of that type and the expected value.

Note: this does not take into account large changes in demand that potentially could impact which patient types are included. This happens if the increase pushes the expected value higher than the price for the lower patient types.

Expected value wrt $p_l$:

$$\frac{\partial EV_{K_j}}{\partial p_l} = \frac{\lambda_k}{\lambda_0 + \sum_{k \in K_j} \lambda_k} = \text{Prob}_{k,j} > 0$$
Note: this does not take into account large changes in price that potentially could impact which patient types are included. This happens if the increase pushes the expected value higher than the price for the lower patient types.

Appendix D: Two private insurers
In this appendix, I work out the solutions for two private insurers (indexed with 1 and 2). The conditions from a bargaining equilibrium are generally:

$$p_{1j} = (1 - \alpha)\Delta WTP_{1j} + \alpha \left[ \sum_{k \in K_j/1} \lambda_k p_k \right] / \left[ \lambda_0 + \sum_{k \in K_j/1} \lambda_k \right]$$

In the two-MCO case, this yields the following system of two equations in two unknowns:

$$p_{1j} = (1 - \alpha)\Delta WTP_{1j} + \alpha \lambda_2 / (\lambda_0 + \lambda_2)p_{2j}$$

$$p_{2j} = (1 - \alpha)\Delta WTP_{2j} + \alpha \lambda_1 / (\lambda_0 + \lambda_1)p_{1j}$$

Which can be rewritten in the following matrix form:

$$A \begin{bmatrix} p_0 \\ p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} 0 \\ (1 - \alpha) WTP_1 \\ (1 - \alpha) WTP_2 \end{bmatrix}$$

Where:

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -\alpha \left( \frac{\lambda_2}{\lambda_0 + \lambda_2} \right) \\ 0 & -\alpha \left( \frac{\lambda_1}{\lambda_0 + \lambda_1} \right) & 1 \end{bmatrix}$$

In this formulation, it is assumed that the provider contracts with all insurers.
This leads to the following equilibrium prices:

\[ p_0 = 0 \]

\[ p^*_1 = \frac{1}{1 - \alpha \left( \frac{\lambda_1}{\lambda_0 + \lambda_1} \right) \alpha \left( \frac{\lambda_2}{\lambda_0 + \lambda_2} \right)} \left[ (1 - \alpha) \text{WTP}_1 + \alpha \left( \frac{\lambda_2}{\lambda_0 + \lambda_2} \right) (1 - \alpha) \text{WTP}_2 \right] \]

\[ p^*_2 = \frac{1}{1 - \alpha \left( \frac{\lambda_2}{\lambda_0 + \lambda_2} \right) \alpha \left( \frac{\lambda_1}{\lambda_0 + \lambda_1} \right)} \left[ (1 - \alpha) \text{WTP}_2 + \alpha \left( \frac{\lambda_1}{\lambda_0 + \lambda_1} \right) (1 - \alpha) \text{WTP}_1 \right] \]

The relationship between prices and size

The equilibrium price ratio is:

\[ \frac{\text{WTP}_1 + \alpha \left( \frac{\lambda_2}{\lambda_0 + \lambda_2} \right) \text{WTP}_2}{\text{WTP}_2 + \alpha \left( \frac{\lambda_1}{\lambda_0 + \lambda_1} \right) \text{WTP}_1} \]

If two insurers have the same WTP then the ratio of prices is:

\[ \frac{p^*_1}{p^*_2} = \frac{\left[ 1 + \alpha \left( \frac{\lambda_2}{\lambda_0 + \lambda_2} \right) \right]}{\left[ 1 + \alpha \left( \frac{\lambda_1}{\lambda_0 + \lambda_1} \right) \right]} \]

\[ \frac{p^*_1}{p^*_2} = \frac{\left( \frac{\lambda_0 + \lambda_2 + \alpha \lambda_2}{\lambda_0 + \lambda_2} \right)}{\left( \frac{\lambda_0 + \lambda_1 + \alpha \lambda_1}{\lambda_0 + \lambda_1} \right)} \]
\[
\frac{\partial}{\partial \lambda_j} \frac{\lambda_j}{\lambda_0 + \lambda_j} = \frac{\lambda_0}{(\lambda_0 + \lambda_j)^2} > 0
\]

Which means that if insurer 1 is larger, the denominator will be larger than the numerator and the ratio will be less than 1, meaning insurer 1 will pay less. The mechanism is that the expected value to the provider without insurer 1 is smaller than the expected value without insurer 2.

Appendix E: Two MCOs and Medicare

In a manner similar to Appendix D, I work out the solutions for two private insurers (indexed with 1 and 2) and an exogenously set public payor (indexed with m). The conditions from a bargaining equilibrium are generally:

\[
p_{ij} = (1 - \alpha)\Delta WTP_{ij} + \alpha \left[ \sum_{k \in K_{ij}} \lambda_k p_k \right] / \left[ \lambda_0 + \sum_{k \in K_{ij}} \lambda_k \right]
\]

\[
p_{1j} = (1 - \alpha)\Delta WTP_{1j} + \alpha \frac{\lambda_2}{\lambda_0 + \lambda_2 + \lambda_m} p_{2j} + \alpha \frac{\lambda_m}{\lambda_0 + \lambda_2 + \lambda_m} p_m
\]

\[
p_{2j} = (1 - \alpha)\Delta WTP_{2j} + \alpha \frac{\lambda_1}{\lambda_0 + \lambda_1 + \lambda_m} p_{1j} + \alpha \frac{\lambda_m}{\lambda_0 + \lambda_1 + \lambda_m} p_m
\]

\[
p_m = \bar{p}_m
\]
Consider the situation with two private insurers (indexed with 1 and 2). This leads to the following matrix formation of the simultaneous equations:

\[
A \begin{bmatrix}
p_1 \\
p_2 \\
p_m
\end{bmatrix} = \begin{bmatrix}
(1 - \alpha) WP_1 \\
(1 - \alpha) WP_2 \\
\bar{p_m}
\end{bmatrix}
\]

Where:

\[
A = \begin{bmatrix}
1 & -\alpha \frac{\lambda_2}{\lambda_0 + \lambda_2 + \lambda_m} & -\alpha \frac{\lambda_m}{\lambda_0 + \lambda_2 + \lambda_m} \\
-\alpha \frac{\lambda_1}{\lambda_0 + \lambda_1 + \lambda_m} & 1 & -\alpha \frac{\lambda_m}{\lambda_0 + \lambda_1 + \lambda_m} \\
-\alpha \frac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_m} & 0 & 1
\end{bmatrix}
\]
In this formulation, it is assumed that the provider contracts with all insurers.

\[
A^{-1} = \frac{1}{\det(A)} \begin{bmatrix}
\frac{1}{\lambda_2} & -1 & 0 \\
\alpha & 1 & 0 \\
\frac{\alpha}{\lambda_0 + \lambda_2 + \lambda_m} & \frac{\alpha}{\lambda_0 + \lambda_1 + \lambda_m} + \alpha^2 & 1 - \alpha^2 \\
\frac{\alpha}{\lambda_0 + \lambda_2 + \lambda_m} & \frac{\alpha}{\lambda_0 + \lambda_1 + \lambda_m} & \frac{\alpha}{\lambda_0 + \lambda_2 + \lambda_m} \\
\frac{\alpha^2}{\lambda_0 + \lambda_2 + \lambda_m} & \frac{\alpha^2}{\lambda_0 + \lambda_1 + \lambda_m} & \frac{\alpha^2}{\lambda_0 + \lambda_2 + \lambda_m}
\end{bmatrix}
\]

\[
\det(A) = 1 + \alpha \frac{\lambda_2}{\lambda_0 + \lambda_2 + \lambda_m}
\]

This leads to the following equilibrium prices:

\[
p_1^* = \left(1 - \alpha^2 \frac{\lambda_1}{\Lambda - \lambda_2} \frac{\lambda_2}{\Lambda - \lambda_2}\right)^{-1} \left[ (1 - \alpha)WT_{P1} + \alpha \frac{\lambda_m}{\Lambda - \lambda_2} p_m + \alpha \frac{\lambda_2}{\Lambda - \lambda_2} \left( (1 - \alpha)WT_{P2} + \alpha \frac{\lambda_m}{\Lambda - \lambda_1} p_m \right) \right]
\]

\[
p_2^* = \left(1 - \alpha^2 \frac{\lambda_2}{\Lambda - \lambda_1} \frac{\lambda_1}{\Lambda - \lambda_2}\right)^{-1} \left[ (1 - \alpha)WT_{P1} + \alpha \frac{\lambda_m}{\Lambda - \lambda_2} p_m + \alpha \frac{\lambda_1}{\Lambda - \lambda_2} \left( (1 - \alpha)WT_{P1} + \alpha \frac{\lambda_m}{\Lambda - \lambda_1} p_m \right) \right]
\]