Physician Market Power and Medical-Care Expenditures*

Abe Dunn and Adam Hale Shapiro†

April 26, 2012

Abstract

We study the degree to which greater physician market power via consolidation leads to higher service prices in the commercially insured medical-care market. We also examine whether these potentially higher service prices translate into different levels of physician service utilization. We find that physicians in more concentrated markets charge higher service prices. However, due to the unique nature of patient cost sharing as well as the incentives of physicians, these higher prices lead to either no change or, in some cases, an expansion of services. This is in contrast to a typical market, where higher prices attributable to consolidation are thought to decrease quantity demanded.

*We would like to thank Kate Ho for her valuable comments and suggestions at the NBER Winter IO Winter meeting. This paper also greatly benefited from discussions with Ben Bridgman, Marco Castaneda, Jim Dana, Ryan Greenaway-McGrevy, Doug Herman, Kyle Hood, Forrest McCluer, Sean Nicholson, Matt Osborne, Marc Rysman, Donald Shapiro, Bob Town, and Shawn Ulrich. We thank seminar participants at the FTC, Northeastern University, The Ohio State University, the Federal Reserve Bank of San Francisco, UC Merced, Claremont-McKenna College, the NBER Winter IO Meeting and the International Industrial Organization Conference. Seidu Dauda and Eli Liebman provided excellent research assistance.

†Bureau of Economic Analysis

1 Introduction

Physicians play a critical role in determining medical-care expenditures. By acting as the patient's health-care consultant, as well as the medical service provider, they can control the quantity of services provided to the patient.[1] Additionally, by flexing their bargaining muscle, they can also potentially raise the fees they charge to insurance carriers. This puts physicians in a unique position of potentially being able to control both the price and the utilization of services—the two components of medical-care expenditures. This control over expenditures is compounded by the possibility that the fee-for-service arrangement between physicians and health insurance carriers may alter the physician's incentives to provide services. Specifically, all else equal, a higher service price (that is, fees) may incentivize the physician to increase service utilization (Hickson et al. [1987], Hemenway et al. [1990], Gruber, Kim, and Mayzlin [1999], Grant [2008], Decker [2009]). In this study, we empirically assess the degree to which greater physician market power via consolidation can raise service prices. We also examine the degree to which these potentially higher service prices may translate into different levels of service utilization. These effects are identified through the large variation in medical-care expenditures observed across markets (Dunn, Shapiro, and Liebman [2011]).

Examining how physicians' market power affects prices and utilization may be an important step in assessing the cause of overall medical-care spending variation. Numerous studies have documented large increases in overall medical-care expenditures over the last few decades, which now account for more than 17 percent of GDP.[2] Growth in medical expenditures has been accompanied by a trend toward consolidation across health care providers.[3] Thus, it is conceivable that long-run trends in consolidation have contributed to the growth in medical-care expenditures. Furthermore, potential concerns have been raised by some industry experts and antitrust authorities that the recent health care reforms enacted in 2010 may accelerate consolidation because they encourage greater cooperation among providers.[4] Understanding how physician market power affects medical-care

[1] See, for example, Sirovich, Gallagher, Wennberg, Fisher [2008].

[2] See Aizcorbe and Nestoriak [2011], Chernew, Hirth, and Cutler [2009], Chernew, Baiker, and Hsu [2010], Cutler and Ly [2011], Dunn, Liebman, Pack and Shapiro [2010].

[3] Gaynor and Haas-Wilson [1999], Smart [2006] and Liebhaber [2007].

[4] Both the Federal Trade Commission (FTC) and Department of Justice (DOJ) have taken different views on the potential impact of recent health care reform. The FTC see consolidation resulting from these reforms as a potential risk that could lead to higher prices, while the DOJ is seen as more receptive to the potential consumer benefits from the proposed reforms (Thomas Catan [2011] "This Takeover Battle Pits Bureaucrat vs. Bureaucrat." *The Wall Street Journal*). Other health economists, interest groups,

spending may give important insights about the potential outcome of this policy reform.

While there has been an extensive line of research regarding hospitals' ability to leverage their market power into higher fees,[5] there has been very little empirical research regarding physicians' bargaining power.[6] Physicians are distinct from hospitals on important dimensions relating to medical-care expenditures. Specifically, the incentives of physicians to affect their own revenue by shifting services provided to patients is distinct from hospitals because hospitals are usually paid on a disease basis.[7] Since physicians are usually paid on a fee-for-service basis, earning additional revenue for every procedure performed, their incentives may be aligned to respond to price changes by shifting the utilization of services.

A major reason for the lack of research regarding physician consolidation has been the dearth of historical granular data covering physician firms. To add fuel to the fire, one must also be able to accurately link physician-firm data to detailed medical-care expenditure information. This study is unique in this regard as we are able to link together a wealth of historical data on physician firms with a comprehensive data set on commercial payments. The physician data are from the SK&A© physician database and include information on the firm size, specialty, and specific geographic coordinates of over 95 percent of physician firms in the United States. This highly detailed data enable us to construct precise physician concentration measures, specific to a particular geographic area. We link these concentration measures with commercial health insurance claims from the MarketScan® Research Database from Thomson Reuters. The data include individual patient health claims for several million privately covered individuals covering thousands of procedures and hundreds of diseases and types of health plans across the entire U.S. The sheer size of this data is a bit daunting, but proves to be important for identification purposes because there is an enormous degree of heterogeneity in types of health service providers, proce-

and politicians have raised some concerns that the new health care law may spur additional consolidation and harm consumers (America's Health Insurance Plans (AHIP) [2011], Berenson, Ginsburg, and Kemper [2010], and "Hearing on Health Care Industry Consolidation" September 2011).

[5]See Noether [1988], Dranove, Shanley, and White [1993], Lynk [1995], and Keeler, Melnick, and Zwanziger [1999], Town and Vistnes [2001], Capps, Dranove, and Satterthwaite [2003].

[6]Research regarding physician market competition has primarily focused on identifying whether or not physicians actually possess monopoly power. As explained by Gaynor [1995], most of these studies have aimed to infer the presence of market power by searching for monopoly rents and supra-normal returns on investment to education (Sloan [1970], Leffler and Lindsay [1980], Burstein and Cromwell [1985].

[7]Indeed, Cutler [1995] finds that the shift in federal payments to hospital from a cost-based (i.e. per service) payment, to a payment that depends on the diagnosis of the patient, had a measurable and positive impact on readmission rates, caused by eliminating the marginal per service payment.

dures, patient ages, diseases, stages of illness, co-morbidities and plan types.[8] Finally, we link together data from HealthLeaders-Interstudy©, which provide comprehensive information on enrollment for health insurance firms. This allows us to create concentration measures of insurance firms. To simplify our analysis and computation burden, we limit our analysis to cardiologists and orthopedists. We believe these two specialties provide a comprehensive look at the physician market since these are two of the largest specialities and cover a wide spectrum of physicians.[9] Cardiologists represent the broad category of internists treating chronic conditions, while orthopedists represent the broad category of surgeons treating more acute conditions.

This paper employs a unique methodology to study competition that is customized to the features of the physician marketplace and the rich data sources available in this study. First, this paper exploits the detailed micro level data to look at the effects of competition on both service price and utilization at the patient level. Second, the precise geographic coordinates of physicians are used to build a consistently defined concentration measure that takes into account the distance and travel time of patients to competing doctors in surrounding areas. We refer to this measure as the "Fixed-Travel-Time Herfindahl-Hirshman Index" (FTHHI). Similar to the measure used in Kessler and McClellan [2000], the FTHHI is designed to remove endogeneity bias stemming from higher quality providers attracting more patients. Third, using individual patient claims and a program provided by Thomson Reuters that categorizes claims into "episodes" of treatment (called the Medstat Episode Grouper®), we are able to build a uniform measure of the quantity of physician services per episode of treatment. This allows us to study the effects of competition on a consistently defined measure of service utilization, which has not previously been studied.

To motivate our empirical analysis, we develop a three period model that outlines how market-structure is linked to the determination of medical-care expenditures. An essential feature of the model is that we divide the components of medical-care expenditures between those variables decided before the patient has sought care and those variables decided after. For example, physician's fees are usually negotiated on an annual basis, and can therefore be considered set before the patient is treated by the physician. In terms of our analysis, we first assess the effect of physician bargaining leverage on service prices, and subsequently

[8]The payment information used to construct prices in this database are the actual negotiated amounts paid to providers and not the "charges"or list price that have been the basis of many prior studies of health care competition.

[9]The Major Practice Categories of "Cardiology" and "Orthopedics & Rheumatology" are the two highest expenditure categories for the commercially insured population (See Aizcorbe and Nestoriak [2011] and Dunn, Liebman, Pack and Shapiro [2010]).

assess the effects of service prices on the utilization of services. For completeness, we also incorporate the first stage impact of insurance carrier market power on both the negotiated service prices and benefits, and the impact of these variables on utilization.

We find that physician concentration is positively and significantly correlated with service price levels. Specifically, a 10 percent increase in the FTHHI will cause about a 1 percent increase in physician fees. Linking this finding to historical survey data discussed in Rebitzer and Vortruba [2011] implies that physician consolidation has caused about an 8 percent increase in fees over the last two decades (1988 to 2008).[10] We also find that health-plan concentration is inversely correlated with service price fees. That is, insurance carriers in more concentrated health insurance markets pay lower fees to physicians.

Having estimated the determination of service prices, we next estimate the effects of the variation in service prices on the utilization decision. A key to our identification stems from our theoretical model, which outlines how patients and physicians likely respond to price variations. Patients with ungenerous benefits will face large variations in their out-pocket expenses due to movements in the service price. By contrast, patients with generous benefits will be less sensitive to movements in the service price, allowing physicians to move along their supply curve. We exploit this dichotomy in order to estimate demand and supply elasticities with respect to service price using a structural switching regression framework. We find a price elasticity of supply in the range of 0.27 to 0.34 for orthopedists and 0.57 to 1.26 for cardiologists. While in most markets an upward sloping supply curve would be unsurprising, in the health service market, this means that physicians treat patients according to service price levels. In other words, a physician with a higher price-cost margin will perform more services. On the demand side, we find a service price elasticity of demand in the range of -0.32 to -0.43 for orthopedics and -0.05 to -0.28 for cardiology patients. This finding supports prior research which suggest that patients are price sensitive, but relatively inelastic (Manning et al. [1987] and Keeler and Rolph [1988]).

Depending on the relative pull between physicians and patients, utilization can either increase or decrease as a result of an increase in physician market power. Our estimates imply higher physician bargaining leverage (and lower insurance carrier bargaining leverage) raises fees, but the effect on utilization depends on the particular market studied. For

[10] Rebitzer and Vortruba [2011] report statistics from a series of physician surveys conducted by the American Medical Association—the proportion of physicians in solo practice, two to four physicians, five to nine, etc. We calculated two chain-linked series of HHIs based on their survey information as well as the SK&A data—one based on the lower bound of the reported firm size bin and another based on the upper bound. The lower bound estimate implied a growth prices of 8.15 percent and the upper bound implied growth in prices of 8.85 percent.

orthopedic patients, the demand response roughly cancels the physician supply response, resulting in no statistically significant change in service utilization. Interestingly, our estimates in the sample of cardiologists imply that the supply response outweighs the demand response, resulting in higher service utilization. Overall, our findings indicate that the unique nature of patient cost sharing and incentives of physicians leads to either no change or, in some cases, an expansion of services. This is in contrast to a typical market, where higher fees attributable to consolidation are thought to decrease quantity demanded.

This paper is organized as follows. Section 2 provides an overview of the physician and insurance carrier industry. We provide a basic framework of physician-insurance carrier bargaining, intended to illustrate how bargaining leverage can affect service prices as well as service utilization. In Section 3, we give a comprehensive overview of the data used in this study. We provide quite a bit of detail as to the construction of our variables as this study includes a battery of new measures not discussed in prior research. In Section 4 we estimate the determination of service prices and benefits and in Section 5 we estimate the determination of service utilization. In Section 6, we quantify the effect of bargaining leverage on service utilization. We conclude in Section 7.

2 Physician and Health Insurance Carrier Organizations

2.1 Physician Organization

The study of physician consolidation has historically been a relatively uninteresting topic due to the fact that a vast majority of physicians worked in solo practices. However, the market for physicians has shifted dramatically over the past few decades. In 1965 only about 10 percent of physicians were in group practices with three or more physicians, but by 1991 group practice physicians accounted for more than 30 percent of all physicians (Smart [2006]). This trend continued through the 1990s and early 2000s. Based on physician surveys, the proportion of physicians in solo practices decreased from 49 percent in 1988 to 33 percent in 2001 (Rebitzer and Vortruba [2011]). There is concern that the recent passage of the health care reforms in 2010 may accelerate the pace of consolidation because the law encourages greater cooperation among providers through the formation of Accountable Care Organizations (ACOs).[11] For instance, a 2011 *New York Times* article by Robert

[11] An ACO is a network of providers that share the provision of care to patients. An ACO would normally include both physicians and hospitals and would encourage greater coordination of care among providers

Pear ("Trade Commission Challenges a Hospital Merger") reports that federal officials are seeing a wave of mergers, in part because of the incentives built into the new health law.

As physician consolidation grew there emerged wider variation in the type of physician practices. Physician group practices vary in size as well as the degree of specialization. Most group practices consist of physicians of the same specialty (that is, single-specialty groups) but there also exist groups with differing specialties (that is, multi-specialty groups). Physician groups can occur as part of a larger health system that contain other group practices, as well as hospitals (that is, Physician-Hospital Organization [PHO]). More complex forms of horizontal structures may involve group practices clustering with one another for bargaining purposes.[12]

Although there are a variety of organizational structures, this paper focuses on the horizontal aspects of these organizations where physicians with the same type of specialization are part of the same group or system. This type of horizontal consolidation has clear implications for bargaining and leverage with health insurance plans.[13] The source of the bargaining power rests on the ability of physicians to threaten to exclude its group from an insurance carrier's network, which may cause significant harm to the profitability of the health insurer. For example, an insurance carrier may find it challenging to attract and adequately treat enrollees if it has only a limited number of cardiologists or orthopedists.

2.2 Insurance Carrier Organization

Similar to the physician market, the health insurance market includes a wide variety of types and sizes of firms (that is, health insurance carriers). They can range in size from small local firms to large firms that are national in scope.[14] Insurance carriers compete

through financial incentives.

[12]For example, two physician groups may have distinct offices and administrative services, but may contract with insurance carriers for legal bargaining purposes as an independent practice association (IPA). In most states, IPAs represent physicians who only compete for *capitated* HMO contracts. For non-capitated contracts, the physicians must negotiate individually unless the FTC rules that they are "clinically integrated" for efficiency reasons (Berenson, Ginsburg, and Kemper [2009]). (Here we focus on specialists that are less often subject to capitated payments where the IPA market stucture is less applicable.) In another example, two physician groups may join forces to share administrative services (e.g. a group practice without walls (GPWW)) as well as contracting.

[13]There have been fewer studies of the effects of consolidation in health care markets along the vertical dimension where the theoretical impact of this type of consolidation is ambiguous (See Cuellar and Gertler [2006] and Ciliberto and Dranove [2006]).

[14]Dranove, Gron, and Mazzeo [2003] show that this type of differentiation can be important in how insurance firms compete.

with one another to attract enrollees. Three important characteristics that differentiate plans in the eyes of the patient are (1) the size of its provider network, (2) restrictions on the patients' choices and (3) the overall price of its insurance contract. Generally, it is assumed that consumers prefer a large choice of providers, less restrictions, and lower prices.

The overall size of the insurance carrier's network is determined according to the bargaining outcome with providers, which we will discuss in the next subsection. Although most commercial health insurance plans have a network of providers, these network insurance plans differ in the restrictiveness of their network.[15] There is a spectrum of types of plans ranging from the least restrictive PPO plans that often contain a broad network of providers and include out-of-network coverage, to the most restrictive type of plans, health maintenance organizations (HMO). Generally, but not always, HMOs will not cover out-of-network providers and will also require a primary-care physician to act as a gatekeeper for seeing a specialist. Finally, the overall price of the insurance contract usually includes the price of the premium, as well as out-of-pocket payments such as a deductible, a coinsurance payment or a fixed co-payment.

Of course, the insurance carrier would like to increase market share, but it would also like to lower its overall costs, which includes the expenditure of treating an episode of care. There are two primary ways in which an insurance carrier can attempt to control the expenditure of an episode of care. It can (1) attempt to lower payments (that is, fees) made to providers or it can (2) attempt to lower the amount of services per episode of care (that is, service utilization). Like the size of the provider network, fees will be bargained over with providers, which we will discuss below. There are a few ways in which an insurance carrier can lower the utilization of services per episode of care. One way is by persuading *patients* in the form of lower benefits. That is, the insurance carrier can pass on some price sensitivity to patients by sharing the cost of the services. A second way is by including more restrictions in the plan, such as monitoring the physicians' actions. For instance, before implementing a procedure, an insurance carrier may require the patient to verify that this procedure is appropriate according to the insurance carrier's medical director. A third, and a bit less intuitive way, is for the insurance carrier to dissuade the *physician* via fee schedules, which works in tandem with how physicians and health insurance carriers bargain over fee payments. We discuss physician-insurance carrier bargaining below.

[15]According to the Kaiser Health Benefit Survey less than 3 percent of enrollees in the U.S. had a conventional indemnity insurance plan.

2.3 Physician-Insurance Carrier Bargaining

Both an insurance carrier's and a physician group's bargaining power resides in the ability to credibly exclude the other side from its patient base.[16] Through a simple framework, we show how each side can use this leverage to affect the expenditure of an episode of care. Specifically, the framework shows how the relative level of concentration on each side of the bargaining table can transmit into variations in service prices (that is, physician fees) as well as service utilization (that is, the quantity of services per episode)—the two components of episode expenditures. To keep the theory parsimonious, we assume symmetric information between patients and these two players.[17]

Assume that each patient pays a fixed premium for an insurance contract that guarantees a minimum number of health "services," $\underline{Q}$, in the occurrence of an episode of illness. Physicians face the cost function $\Psi(Q)$ and marginal cost $\psi(Q)$, where the physician's marginal cost will be an increasing function of the number of services provided (that is, $\psi'(Q) > 0$).[18] One can think of an episode of care as the time period between the initial health shock to an individual and final treatment.[19] The total expenditure of treating an episode of care, TE, is thus the service price, P, times the total number of services, Q (that is, $TE = P \cdot Q$).[20] Letting α represent the proportion of expenditures paid by the patient (a measure of the generosity of benefits), the patient's demand for services can be represented as $D(P^{pock})$, where $P^{pock} = \alpha P$ is the out-of-pocket price.

The determination of these variables is easier to understand if one thinks of them occurring in three distinct periods:

[16] Staten, Umbeck, and Dunkelberg [1987, 1988] as well as Sorenson [2003] are studies showing that an insurance carrier's bargaining leverage resides in its credibility of exclusion.

[17] Full information is an overly strong assumption in the market for healthcare, however, it allows us to isolate the effects of the competitive bargaining game between the insurance carrier and the physician.

[18] There are several factors that may cause physician firms to have increasing marginal costs for the treatment of patients. Perhaps the most important factor is the opportunity cost of the physician's time. The physician's limited amount of time in a day will make it necessary to hire additional units of labor or capital as she expands the number of services provided per patient (e.g. assistants or other physicians may be added to the physician firm) which may be costly. It is also possible that physicians perceive that the probability of malpractice lawsuits or damage to their reputation are higher as more services are done that may be less beneficial to the patient. In any case, the empirical model in this paper will test this assumption. Note that if the marginal cost curve is flat, we should not expect to find an empirical relationship between the physician's price and utilization.

[19] In our empirical analysis, we will cap an episode for a chronic disease at 365 days.

[20] TE is sometimes referred to as the "episode price" in the literature.

Period 0: Entry of physician and insurance firms.

Period 1: *(a)* Service price (P) negotiated.

(b) Insurance benefits (α) chosen.

Period 2: Service utilization (Q) decided.

One can think of Period 0 as the time period when the long-run equilibrium entry of insurers and physicians occurs, which determines market structure. Period 1 and Period 2 can be thought of as a dichotomy between those variables determined before the patient seeks care versus those variables determined after the patient seeks care. Specifically, because fee negotiations and insurance plan choices usually take place on an annual basis they can be thought of as being determined prior to when the patient seeks care. In Period 1, the physician and the insurance carrier bargain over service prices, defined as the price per service paid to the physician by the insurance carrier, and the insurance carrier and the patient determine a benefits package for the patient. In Period 2, an optimal service utilization decision is made.

2.3.1 Period 0

In Period 0 both insurance carriers and physician firms consider the profitability of entering different local markets. In the spirit of Bresnahan and Reiss (1991), insurance carriers and physician firms enter a market if the expected profits from entry are positive. The key components of profits include the variable profit per individual served in the market, the market size, along with a fixed entry cost. More formally, an insurance carrier enters a local market if its profit from entry is positive:

$$\pi_{ins} = M \cdot \mathbf{d}(prem, P^{pock}) \cdot [prem - \mathbf{AVC}(P^{pock}, P, Q(P^{pock}, P))] - F > 0 \tag{1}$$

where $\mathbf{d}(prem, P^{pock})$ represents individual insurance demand and is a function of both the premium, $prem$, and the amount paid out-of-pocket, P^{pock}. The market size covered by the insurance carrier is represented as M. The average variable cost of the insurance carrier, $\mathbf{AVC}(P^{pock}, P, Q(P^{pock}, P))$, is a function of utilization, the service price, and the out-of-pocket price. The insurance carrier's fixed cost is F.

The physician firm's entry decision is also determined by its profit function. The physician firm enters if profits are positive:

$$\pi_{phys} = m \cdot \mathbf{s}(P) \cdot [\mathbf{TE}(P, P^{pock}) - \mathbf{\Psi}(Q(P^{pock}, P))] - f > 0. \tag{2}$$

where total physician revenues are represented as $\mathbf{TE}(P, P^{pock}) = P \cdot Q(P^{pock}, P)$, physician cost, $\mathbf{\Psi}(Q(P, P^{pock}))$, is a function of utilization and the physician firm's market share is a function of the service price, $\mathbf{s}(P)$. Here, m is market size covered by the physician—the number of individuals in the physician's geographic market that are expected to have an episode. The profit per patient is the price per service, P, times the amount of services per episode, Q, minus the total cost of those services Ψ. The physician firm's fixed cost is f.[21]

A key point to take away from this section is that there are distinct factors that affect physician and insurer entry decisions. Namely, the size of the market facing insurance carrier, M, is distinct from the market facing the physician firm, m. Physicians tend to operate in relatively local geographic markets and provide physician services to individuals across a wide range of insurance types, including commercial market enrollees, Medicare enrollees, Medicaid enrollees, and individuals enrolled in COBRA. In contrast, health insurance carriers typically offer insurance to employers over a broader geographic area, such as an MSA. These insurance carriers disproportionately serve individuals that are not in the government-funded programs. Insurance carriers also sell their services to a wide variety of employers, which may have an impact on the amount of revenue that they receive and how they operate their business. While larger firms typically offer health insurance to their employees and tend to purchase self insurance, smaller employers are much less likely to offer insurance and they tend to purchase full insurance. In addition, the fixed costs, F and f, associated with entering a new market are likely very different for physicians and insurance companies since they offer distinct services.

2.3.2 Period 1

In Period 1, the level of benefits is set and the service price is negotiated. A major determinant of the level of benefits purchased, α^*, will be the level of concentration in the insurance carrier market (for example, Dafny et al. [2011] and Dunn [2011])—where the star superscript indicates it was decided upon in Period 1. Specifically, if an insurance carrier has a larger degree of market power, it will be in its interest to offer less generous benefits (that is, a larger α). For simplicity, we assume that during fee negotiations physicians expect that all patients have chosen the same benefit structure.

[21]Note that the profit functions presented here have been simplified considerably for expositional purpose, with each physician having only a single price and single utilization level per episode. In reality, physicians treat a very heterogenous population of individuals with a variety of insurance types and health conditions, and this more detailed information will be incorporated into our empirical framework.

To help motivate the empirical framework and explain the determinants of the negotiated service price, P^*, we present a stylized bargaining model. In this model, the service price (that is, fee schedule) will depend on negotiations between the physician group and the insurance carrier. Although fixed costs affect entry, subsequent to Period 0, the fixed costs are sunk and are no longer important in the determination of the service price. During fee negotiations, both the insurance carrier and the physician firm will be concerned with the potential set of enrollees that the physician firm would cover. The negotiation with the physician firm will affect each insurance carrier's profits in multiple ways, but here we emphasize the effects of the negotiated service price, P^*, on the average cost of the insurer serving the m potential patients. Let the insurer's variable profit function in Period 1 be:

$$\pi_{ins}^{V} = -\mathbf{AVC}(P^{pock}, P, Q(P^{pock}, P)) = -[1-\alpha] \cdot \mathbf{s}(P) \cdot \mathbf{TE}(P, P^{pock}) \tag{3}$$

The physician's variable profit function in Period 1 is:

$$\pi_{phys}^{V} = \mathbf{s}(P) \cdot [\mathbf{TE}(P, P^{pock}) - \mathbf{\Psi}(Q(P, P^{pock}))]. \tag{4}$$

It is imperative to highlight that the measures of the potential market size, m and M, affect profitability; however, they are unlikely to have a direct effect on the negotiations between insurers and physicians, *except* through an impact on the competitive environment—determined in Period 0. To see this important concept, suppose the negotiated service price is determined by a specific bilateral Nash Bargaining problem (first proposed by Horn and Wolinsky [1988]). For simplicity, assume that the physician firm's geographic market lies within that of the insurance carrier's geographic market. It follows that the overlapping market size of the insurance carrier and the physician firm is simply m. Then, the expected impact on the profit from this population is $m \cdot \pi_{ins}^{V}$ for insurers and $m \cdot \pi_{phys}^{V}$ for physicians. In this case, each bilateral price maximizes the Nash Product of the insurer and physician profits:

$$\max_{P} \left[m \cdot \pi_{ins}^{V} - m \cdot \delta_{ins}\right]^{b_{ins}} \left[m \cdot \pi_{phys}^{V} - m \cdot \delta_{phys}\right]^{b_{phys}}$$

where δ_i is the disagreement payoff for either the physician firm or the insurance carrier. One may think of δ_{ins} as the expected AVC for a patient who is treated by the outside option (for example, the cost of a patient seeing another physician that does have a contract with the insurer). It follows that the total insurance carrier disagreement payoff is $m \cdot \delta_{ins}$. On the physician side, δ_{phys} represents the expected variable profit to the physician for treating a patient insured by the outside option (for example, the variable profit from treating patients insured by carriers that do have a contract with the physician). Similarly, the total physician disagreement payoff is $m \cdot \delta_{phys}$. Note that since both payment and disagreement amounts are proportional to m, we may write the maximization problem as

$$\max_P [m^{b_{ins}} m^{b_{phys}}] \left[\pi^V_{ins} - \delta_{ins}\right]^{b_{ins}} \left[\pi^V_{phys} - \delta_{phys}\right]^{b_{phys}}$$

Considering the first order conditions to this maximization problem, one can see that the market size information would drop out of the equation during Period 1 negotiations, so the maximization problem may be re-written as:

$$\max_P \left[\pi^V_{ins} - \delta_{ins}\right]^{b_{ins}} \left[\pi^V_{phys} - \delta_{phys}\right]^{b_{phys}} \tag{5}$$

Therefore, in this bargaining game, the market size and fixed costs do not directly enter this first period profit maximization problem. However, m, M, f, and F clearly affect entry in Period 0, which will subsequently have an impact on this maximization problem and the negotiated price. That is, pricing decisions in the market will be related to market size and fixed costs, *only* through the effects on the competitive environment.[22]

Although this stylized bargaining model provides some intuition for the determinants of the negotiated service price, there are several reasons why we empirically analyze the reduced-form relationship between the competitive environment and negotiated prices. First, the above bargaining model is highly simplified and does not reflect the complexity of the actual bargaining environment. For instance, it is relatively common for bargaining between insurers and providers to break down, so there is no negotiated contract; and the possibility that negotiations may fail, can result in multiple equilibria in the insurer and provider contracting decisions (for example, see Ho [2009]). Therefore, it is not clear what pricing game is appropriate to consider in this market. Second, in this paper we are less interested in identifying the underlying structural parameters and profit functions of insurers and physicians, and more interested in understanding the relationship between the competitive environment and the observed outcomes, as in Davis [2005, 2006a]. Finally, although we are using extremely rich and geographically diverse data, there are limitations in the data, as we will describe later, that make a more structural analysis challenging.

[22] A similar argument is relevant for the insurer when setting premiums and benefits, P^{pock}. Specifically, the insurer solves the maximization problem $\max_P M \cdot \mathbf{d}(\phi, P^{pock})(\phi - \mathbf{AVC}(P^{pock}, P, Q(P^{pock}, P)))$. In this optimization problem, the market size relevant to the insurer, M, and the fixed cost, F, also drop out of the first order condition.

Instead, we develop an empirical framework where the service price will depend on the relative degree of bargaining power, we label Z, of the physician and the insurance carrier in the market. One possible measurement of Z is the logarithm of the ratio of concentration measures: $Z = \ln\left(\frac{HHI_{phys}}{HHI_{plan}}\right)$, where HHI_{phys} and HHI_{ins} are the degrees of concentration in the physician and insurance carrier markets, respectively. These concentration ratios reflect the strength of the outside options of physicians (when considering contracts with different insurers) and insurers (when considering contracting with different physicians). To examine how service price may depend on Z, we look at two polar market structures:

$$\begin{aligned} Z_1 &\approx -\infty \quad \Rightarrow \quad \text{Competitive Physician, Monopolistic Insurance Carrier.} \\ Z_2 &\approx \quad \infty \quad \Rightarrow \quad \text{Monopolistic Physician, Competitive Insurance Carrier.} \end{aligned}$$

Moving from market structure Z_1 to Z_2, we are shifting market power leverage from insurance carriers to physicians. Under market structure Z_1 the monopolistic insurance carrier can credibly threaten to keep the competitive physician out of its network. This credible threat will subsequently induce physicians to bid the price of services down to to the point where $P = \psi(\underline{Q})$, the physician's marginal cost of providing the minimum amount of services. By contrast, under market structure Z_2, the monopolistic physician can credibly threaten to exclude the insurance carrier's patients from using its services. Specifically, provided that the risk-free rate of return is earned, there will always be at least one insurance carrier willing to accept the profit-maximizing monopoly price.[23] Thus, any insurance carrier who wants to contract with this physician must offer the profit maximizing service price.

The actual marketplace is rarely perfectly competitive or completely monopolistic. Instead, prices will be pulled towards either of these two extremes by the side with larger bargaining leverage. Thus, bargaining leverage is manifested in price variations by each side's ability to credibly exclude the other from its network. It is also important to note that this type of bargaining is usually implicit, rather than direct, interactive bargaining. That is, a health care plan may not directly discuss or haggle with a physician firm over price, but rather just recognize its relative competitive position and create a payment schedule that would entice the physician firm to participate in the plan.

[23] That is, the monopoly price will be $P = \frac{\varepsilon}{1-\varepsilon}\psi^{mon}$, where ε is the patient's demand elasticity and ψ^{mon} is the marginal cost at the monopoly quantity of services.

2.3.3 Period 2

After benefits and service prices are set in Period 1, the utilization of services is determined. Under certain simplifications, the physician's profit-maximizing amount of services to provide to the patient is quite intuitive. For example, let us assume that out-of-pocket costs are small (that is, α is close to zero) such that patients are not sensitive to variations in the service price. Let n represent the patient-episode-physician triple. It follows that the physician solves the following:

$$\max_{Q}\ Q_n P_n^* - \int_0^{Q_n} \psi(t)dt \tag{6}$$

where P_n^* is the service price set in Period 1. This results in the profit maximizing number of services:

$$Q_n^* = \psi^{-1}(P_n^*) \tag{7}$$

Given a Period 1 negotiated service price, the physician will provide the quantity of services to the patient up to the point where his marginal revenue equals his marginal cost. It follows that as long as marginal cost is increasing with the number of services, (that is, $\psi'(Q) > 0$), then the physician's optimal utilization is *increasing* in the pre-negotiated service price.[24] In other words, in the second-period the physician acts as a price-taking firm with the traditional, upward-sloping supply curve.

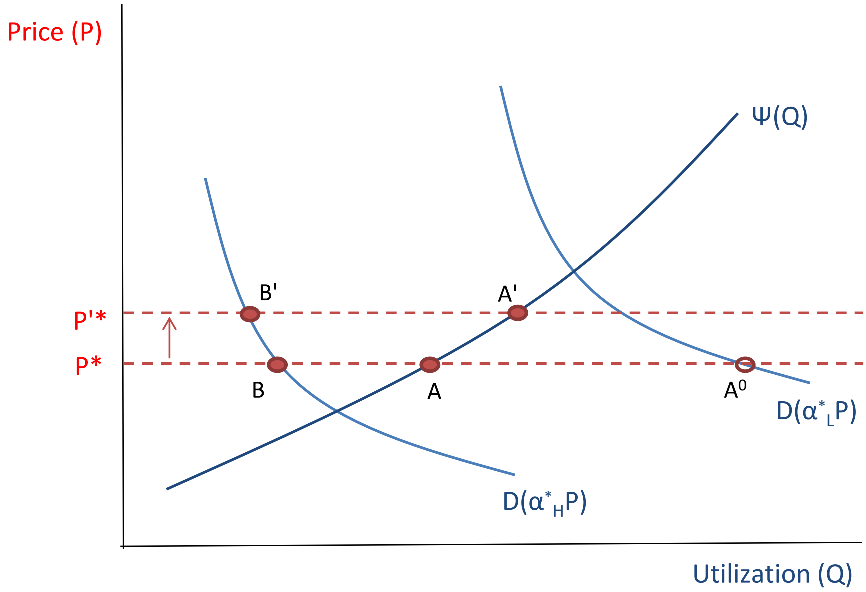
Under more general conditions, there will exist a distribution of patients, each having distinct benefit schedules, α_n^*. In this case, the physician's profit maximization problem is identical to (6), except that it includes the constraint that the patient perceives that each additional service provides a positive net marginal benefit.[25] This constraint implies that the physician cannot force the patient to consume more services than the patient demands.[26] The key insight here is that variations in the service price will affect the patient's net marginal benefit of receiving services depending on the level of α_n. A less generous benefit schedule (that is, a larger α_n) will cause the patient to become more sensitive to variations in the negotiated service price.

[24] This follows since $\frac{d\psi}{dQ}^{-1} = \frac{1}{(d\psi/dR)} > 0$ if $d\psi/dQ > 0$.

[25] More formally, if the patient faces an out-of-pocket price of, P_n^{pock}, then the patient's demand for services is $D(P_n^{pock})$. In the case where $D(P_n^{pock}) \leq \psi^{-1}(P_n)$, then the patient's demand constraint is binding and $Q_n = D(P_n^{pock})$. Alternatively, in the case where $\psi^{-1}(P_n) < D(P_n^{pock})$, then the physician's profit function is binding, so that $Q_n = \psi^{-1}(P_n)$.

[26] This is identical to the assumption in Dranove [1988] where he assumes that "The patient receives treatment if and only if the physician recommends it and he consents."

Figure 1: Period 2 Utilization Decision



This concept can be demonstrated by examining Figure 1 where we have a diagram depicting the second-period service utilization decision. Here, the first-period negotiated price is denoted as the constant dashed line at P^* and the marginal cost curve is denoted by the upward sloping curve, $\psi(Q)$. We have also depicted two different demand curves, corresponding to patients with benefit schedules α_L^* and α_H^* where $\alpha_L^* < \alpha_H^*$.[27]

Under benefit structure α_L, the patient faces a low out-of-pocket price and demands services corresponding to point A^0. At this level of services, the service price does not cover marginal cost, which means the physician can only provide services up to the point A, where service price equals marginal cost. Since the utilization decision resides on the physicians marginal cost curve, the patient's demand curve is not binding and a marginal increase in the first-stage negotiated price to P'^* will increase the utilization of services to A'. This is the outcome shown in equation (7) where we assumed that the patient faced low out-of-pocket costs, $\alpha \approx 0$. By contrast, let us examine the scenario where the patient faces a large out-of-pocket price, $P^{pock} = \alpha_H^* P^*$. Specifically, at a first-period negotiated service price of P^*, the patient faces an out-of-pocket price of $\alpha_H^* P^*$, and will demand services at point B. As the service price at this level of utilization covers the physician's marginal cost, the demand curve is binding. In this case, a marginal increase in the first-

[27]It follows that as $\alpha_n \to 0$, the price paid by the patient falls for any given P^*, and the demand curve shifts to the right.

period negotiated price (from P^* to P'^*) corresponds to a movement along the demand curve and will consequently lower the utilization of services provided.

The overall effect of a change in service price on service utilization will therefore depend on the amount of benefits provided to the patient. When benefits are low, such that the out-of-pocket price is relatively high, utilization is likely to be negatively correlated with service prices. However, when benefits are high, such that out-of-pocket costs are low, a higher service price may ultimately raise the quantity of services provided. Thus, insurance carriers can thwart any effect of a positive supply elasticity by raising the amount of cost sharing, α_n, to the patient. Overall, this framework is intended to show how market power can dictate physician episode expenditures, TE_n. First, larger physician bargaining leverage can translate into higher fees P_n. Second, if large proportion of patients are receiving generous benefits, such that they do not face the full service price, these higher prices may result in more services provided to the patient.

Discussion The ability of physicians to affect the utilization of services has been modeled theoretically by a number of researchers studying supplier induced demand (for example, Evans [1974], Fuchs [1978)], Dranove [1988], and McGuire and Pauly [1991]), but much of the literature relies on the assumption that physicians may recommend treatment that the patient would not have desired under symmetric information. This is a controversial assumption in both the economic and medical fields, since it suggests that physicians may push unnecessary services onto patients. Some advantages of the framework presented here are that it avoids this assumption. It shows that "physician inducement" can arise even without asymmetric information and it adheres to a more neoclassical theory where prices and quantity are determined by market forces and preferences of consumers are fixed. Just like in a typical market, firms are profit maximizing and all services purchased by the consumer provide a perceived net benefit. In fact, in the framework presented here, it is often the case that physicians are actually constraining the number of services used by the patient (that is, the patient would be willing to undergo additional treatment upon recommendation).[28]

Another advantage of this model is that it combines the price responses of both the physician *and* the patient into a unified framework. Specifically, the model allows for both a positive supply response from physicians as well as a negative demand response by patients.

[28] It is not clear that restraining the number of services is actually harmful to the patient in the first period. Due to the classic moral hazard issues, the additional service utilization in the second period may lead to a welfare loss from higher premiums that are necessary to cover greater expected medical expenditures.

In previous literature, these effects have been modeled separately. It is important to point out that our empirical specification will not constrain the relationship between service prices and utilization in any direction. Thus, we allow for a possible negative relationship as suggested by the theories of Evans [1974], Fuchs [1978], and Dranove [1988].[29]

3 Data

In this section, we give a comprehensive overview of the datasets used in this study. First, we describe the MarketScan® health claims database, which is a database that tracks claims from all providers using a nationwide convenience sample of patients. We also provide an overview of how we calculated our service price and service utilization variables, which we show are components of total physician expenditures. Second, we describe the SK&A© physician database, which includes information on location, specialty, unique physician identifiers, medical practice group, and health system of physicians in the United States. We then give an overview of the HealthLeaders-InterStudy© as well as the Area Resource File data, which provide information used to make concentration measures of health insurance firms as well as demographic information.

3.1 MarketScan® Data

The MarketScan® database tracks claims from all providers using a nationwide convenience sample of patients. Our collected data span 2005 through 2008. The data include health claims from employers and insurance carriers throughout the entire United States; all claims have been paid and adjudicated. Each enrollee and provider has a unique identifier and can be identified at the county level. This paper uses the Commercial Claims and Encounters Database portion of the MarketScan® Databases, which includes healthcare

[29] It is also possible that the alternative inducement theory of McGuire and Pauly [1991] may be the true underlying mechanism through which service prices affect utilization, which also suggests that higher prices can lead to greater utilization in some cases. Consistent with the supplier induced demand hypothesis, their model also suggests the possibility that higher service prices may lead to less utilization if higher service prices increase income by a sufficient amount, so that income effects dominate the substitution effect. Here we do not consider physician "income effects" in our model, since we treat physicians as firms that are not constrained by diminishing marginal utility from leisure. Although income effects are possible and could potentially be incorporated in our model, the consolidation of the physician markets over the past several decades suggests that in today's market viewing physicians as firms may be the more plausible economic assumption.

utilization and cost records at the encounter level. This portion of the database provides patient identifiers that may be used to sum expenditures to the patient-episode level.

The Commercial Claims and Encounters Database contains data from employer and insurance carrier sources concerning medical and drug data for several million employer-sponsored insurance (ESI)-covered individuals, including employees, their spouses, and dependents. Each observation in the data corresponds to a line item in an "explanation of benefits" form; therefore each claim can consist of many services and each encounter can consist of many claims.

Importantly we can differentiate between payments made to physicians from those paid to other providers (for example, hospitals and pharmacies). For instance, suppose a patient is being treated for congestive heart failure in a hospital. The claims data differentiates between types of providers such that payments made to the physician for performing a coronary artery bypass are distinct from those made for hospital operating room expenses. We use MarketScan's "payment" variable which is defined as the total gross payment to a provider for a specific service. Specifically, this is the amount of dollars eligible for payment after applying pricing guidelines such as fee schedules and discounts, and before applying deductibles and co-payments. MarketScan® also indicates the type of plan the claim was made under, which allows us to ignore episodes in which a capitation payment was made.[30]

3.1.1 Physician Expenditure of an Episode of Care

To obtain the physician expenditure for a particular episode of care we apply the Medstat Episode Grouper® (MEG). This algorithm, provided by Thomson Reuters, assigns a procedure to an episode using information on claims as well as the patient's medical history. Spending is allocated to a patient between a beginning and an end date by assigning an "episode ID", n, to each claim in the data.[31] Let Γ_n be the set of procedures used for treating episode n identified by the MEG. The total expenditures made to the physician for treating episode n is:

$$TE_n = \sum_{j \in \Gamma_n} p_{jn} \tag{8}$$

where p_{jn} is the full payment (including the patient's out-of-pocket costs) to the physician

[30] Approximately 3 percent of our sample are capitated episodes. These observations are likely to include closed HMO systems such as Kaiser-Permanente patients.

[31] We isolated episodes where the patient sees the same physician for the entire episode of care, however, results were not sensitive to this exclusion.

for procedure j in episode n.[32] Pricing information for a specific procedure is the payment attached to the specific health claim line in the MarketScan® data. We identify procedures j at the most granular level possible, based on a specific CPT code, modifier, and "place of service."[33] Note that each episode uniquely identifies an individual patient, k, with disease d, treated by a physician p, in county c, that begins in time period t.[34] The large advantage of the MEG algorithm is that it allows us to isolate the service mix and total price for treating a particular patient's illness. However, these algorithms are also considered a "black box" in the sense that they rely entirely on the expertise of those that developed the grouper software.

3.1.2 Decomposing the Expenditure of an Episode of Care

As outlined in Section 2, the outcome of a bargaining game between physicians and insurance carriers will result in variation in both service prices (that is, the fee schedule) as well as the utilization mix of services in a given episode of care. Thus, embedded in the expenditure of an episode of care is a "service-price component" and a "service-utilization component."

The service-utilization component is the number of services provided to the patient over the course of the episode of care. We measure this variable by the following:

$$Q_n = \sum_{j \in \Gamma_n} r_j \tag{9}$$

where r_j is the average price of procedure j in the entire sample. Here, r_j serves as a proxy for the number of services rendered for each given procedure, and thus one can think of r_j as being comparable to each procedure's relative value units (RVUs) assigned by Medicare. Any variation in the utilization component between two episodes of care will be attributable to differences in the number of services used as opposed to differences in the prices charged for the same service. The remaining component of the expenditure of an episode of care is the service price:

[32]Note that each episode occurs only once in the data, thus we do not have a panel of episodes.

[33]We chose to differentiate procedures by place of service based on the fact that Medicare provides higher fees for physicians who have their own office-based facility.

[34]An episode of care may span several time periods (half-years in our analysis) for chronic diseases. We assign the episode to the date at which the episode begins. For our analysis, we isolated episodes treated by only one physician.

$$P_n = \frac{\sum_{j \in \Gamma_n} p_{jn}}{\sum_{j \in \Gamma_n} r_j} \tag{10}$$

which is the price of the episode of care in terms of its total price per service. In our empirical analysis, we will assess how market power affects each of these components individually. Specifically, we use the fact that in logs our decomposition of total episode cost takes the tractable form:

$$\ln(TE_n) = \ln(Q_n) + \ln(P_n) \tag{11}$$

This equation shows that any percentage change in total physician expenditures, TE_n, will be due to either a percentage change in service utilization or a percentage change in service price.

Table 1 shows summary statistics of P_n, Q_n, and TE_n for our entire MarketScan® sample. On average, cardiologists bill 459 dollars per episode of care while orthopedists bill 463 dollars.[35] Note that due to how we defined services, the mean and median price per service will be approximately equal to one by construction. Overall, the data show a large amount of variation in both prices and utilization. The 90th percentile service price is about twice as large as the 10th percentile service price for both cardiology and orthopedists. There is much wider dispersion in utilization rates, especially for cardiology. The 90th-percentile utilization is 58 times the magnitude of the 20th-percentile utilization for cardiologists and 17 times the magnitude for orthopedics. Although these differences in utilization rates appear large, it is important to note that this variation may partly be explained by the wide variety of different diseases treated by each specialty.

3.2 SK&A© Data

The SK&A© database includes information on physician location, specialty, name, medical practice group, and health system. The database is updated every six months, spans 2005 to 2008, and includes 95 percent of office-based physicians practicing in the United States.[36] One major advantage of the SK&A data over other databases is that each physician is verified over the telephone, which increases the accuracy of its physician

[35]We removed outliers we believe are attributable to clerical data input error by discarding episodes in the bottom first percentile and top 99th percentile based on price per service and utilization.

[36]SK&A has a research center that verifies every field of every record in its database. The data also includes the names of DOs, NPs, PAs and office managers.

Table 1: Summary Statistics

	Mean	SD	10th Ptile	90th Ptile
	Cardiology			
$FTHHI_{phys}$	0.119	0.107	0.026	0.241
HHI_{ins}	0.231	0.100	0.130	0.358
P_n	1.012	0.339	0.684	1.338
Q_n	459.6	717.9	23.0	1336.9
TE_n	459.9	755.1	24.3	1298.7
HMO_n	0.107	0.309	0	1
PPO_n	0.657	0.474	0	1
POS_n	0.106	0.308	0	1
$HDHP_n$	0.002	0.041	0	0
$CDHP_n$	0.019	0.139	0	0
$BMCOMP_n$	0.107	0.310	0	0
$EMPLOYER_n$	0.606	0.488	0	1
AGE_n	51	11	36	63
$COMORBID_n$	5.93	3.37	2	10
α_n	0.27	0.31	0	0.85
Stage of Illness	1.23	0.651	1	2
	Orthopedics			
$FTHHI_{phys}$	0.104	0.105	0.022	0.218
HHI_{ins}	0.233	0.101	0.130	0.358
P_n	1.032	0.279	0.742	1.356
Q_n	462.5	729.7	59.3	1176.3
TE_n	463.7	769.4	67.0	1166.7
HMO_n	0.112	0.315	0	1
PPO_n	0.661	0.473	0	1
POS_n	0.121	0.326	0	1
$HDHP_n$	0.002	0.045	0	0
$CDHP_n$	0.024	0.15	0	0
$BMCOMP_n$	0.079	0.264	0	0
$EMPLOYER_n$	0.576	0.494	0	1
AGE_n	40	18	13	61
$COMORBID_n$	5.59	3.20	2	10
α_n	0.30	0.29	0.025	0.83
Stage of Illness	1.01	0.167	1	1

location and group size information.[37]

Given the different types of physician organizations, assigning each physician to a specific firm is not a straightforward task. One difficulty is how to overcome the complexity in the vertical dimension. For instance, a small portion of our sample (6 percent) of those physicians who belonged to a group medical practice also belonged to a larger health system. Anecdotal evidence from physicians leads us to believe that bargaining in this case would take place at the larger health system level; therefore we make the assumption that physicians use their full market power whenever possible.[38] Thus, for each physician we assign the broadest medical group or system she was assigned in the data. Specifically, if the physician is not associated with a health system we assign her to the group medical practice she is listed with.

Pinpointing the geographic market for provider services is also a challenging task, and has been the subject of many antitrust cases (see Gaynor and Haas-Wilson [1999] and Haas-Wilson [2003]). Neither the Justice Department nor the Federal Trade Commission have a set standard as to how to measure the size of a geographic market for medical services.[39] In creating our measure of the geographic market, we use as much of the granularity of the physician-location information as possible. We define a geographic region as the area surrounding a given patient, as would be done in a standard Hotelling problem. Specifically, for each location in the SK&A data, we create a distinct concentration measure based on the physicians in the surrounding geographic region. To do this takes a few steps. First, we define a geographic region by specifying a maximum amount of driving time, $\bar{k}$, a patient would be willing to travel to see a physician.[40] The value of $\bar{k}$ is 80 minutes and is found by searching for the value that resulted in the lowest mean-squared error in our regression analysis. Second, we calculate the probability that a patient located at the center of the

[37]The six month frequency of their telephone survey may be important, since SK&A reports that on average, 14.2% of physicians move each year. Although all the information in the survey is telephone verified, they gather information for physicians through a variety of sources. This includes company and corporate directories, websites, state licensing information, mergers and acquisitions announcements, trade publications, white and yellow pages directories, professional associations, and government agencies.

[38]This is consistent with the common assumption made in the hospital literature that the hospitals bargain at the system level.

[39]Although many experts agree that the merger guidelines provide an appropriate framework for defining and analyzing geographic markets in the health care sector, there is no consensus for the precise methodology that should be used across all markets (See FTC and DOJ [2004]).

[40]Driving times were calculated in Stata using the 'traveltime' command developed by Ozimek and Miles. This command finds the driving time between two latitude and longitude points via Google maps. See Appendix A for more details.

geographic coordinate would travel to see a physician. We do so based on the assumption of linear travel costs and uniform taste preferences. Third, using the probabilities of seeing each physician calculated in step two, we calculate expected market shares based on the physician group's size and distance to the specific geographic coordinate. In this fashion, those physicians closer to the patient are given more weight than those physicians farther away. Using these market shares, we construct the concentration measure for each coordinate, representing the competition for that patient in the surrounding area, $HHI^{patient}$. Fourth, we link these measures to the MarketScan® data by averaging $HHI^{patient}$ over the county, so that there is one HHI_c specific to a county. Fifth, to arrive at a HHI specific to a physician (the fixed-travel time HHI, $FTHHI_{phys}$), we weight the aggregate county HHI_c measures using information on the county of the patients for each provider in the MarketScan® data. We treated each specialty as distinct from each other, meaning that cardiologists were not counted in the $FTHHI_{phys}$ created for orthopedists and vice versa. More explicit details of the construction of $FTHHI_{phys}$ are available in Appendix A.

It is important to note that for hospitals it is possible to define the market based on a demand estimate using a discrete-choice framework where patients choose among a discrete set of hospitals (see Town and Vistnes [2001] and Capps, Dranove, and Satterthwaite [2003]). However, this paper takes a more reduced form approach for three reasons. First, the discrete-choice framework applied in the hospital literature is not possible with our data because we do not have geographic information in the MarketScan® data at the zip code level (that is, MarketScan® tracks providers and patients at the county level). Second, the number of physician firms is magnitudes larger than the number of hospitals, which means the number of possible physician choices becomes quite large. Third, the effects of competition among physicians are not well understood or documented, so as a first step in analyzing this market we focus more directly on the relationship between the competitive environment and its effects on service prices and outcomes. For these reasons, this paper more closely follows papers that apply more reduced form techniques (for example, Lynk [1995], Dranove and Ludwick [1999], Kessler and McClellan [2000], Duggan [2002], and Dranove et al. [2008]).

Table 1 provides summary statistics for the physician concentration measure, $FTHHI_{phys}$, for both our cardiologists and orthopedists sample. The orthopedist market is slightly less concentrated with an average $FTHHI_{phys}$ of 0.104 versus an average of 0.119 for cardiologists. There is also a wide degree of variation in this variable, as, in both samples, the 90th-percentile measure is roughly ten times larger than the 10th-percentile measure. There is not a large degree of time series variation in the physician HHI variables. The mean cardiology $FTHHI_{phys}$ is 0.123 in the first half of 2005 and is 0.108 in the second half of 2008. The corresponding measures for orthopedists are 0.109 and 0.097.

3.3 HealthLeaders-InterStudy© Data

Enrollment information on health insurers is obtained from the HealthLeaders-InterStudy© database of insurance carriers for the years 2005 to 2008. This MSA level enrollment data are collected through a biannual survey of health insurance carriers where they are asked to report enrollment by geographic location. The enrollment information for each insurance carrier is also provided by the type of health insurance plan (that is, PPO, POS and HMO)[41] and also whether the contract is fully-insured or self-insured.[42]

Using this enrollment data, we construct an HHI concentration measure for the health insurance market. The HHI measure is constructed based on the share of total enrollment for each plan. Specifically, we let S_{ins} be the share of enrollment for an insurance carrier in an MSA, then the concentration measure for the enrollee is $HHI_{ins} = \sum_{ins \in MSA}(S_{ins})^2$.[43]

3.4 Demographic Data - Area Resource File & Census Data

For additional information regarding the demographic information in a county area we use data from the Area Resource File (ARF). The ARF is a database containing extensive

[41] Prior to 2004 HealthLeaders-Interstudy© collected data on only HMOs, but they significantly expanded the coverage of their plan survey in 2004. Prior to 2006 they did not separately report POS, but included this enrollment as part of the HMO category.

[42] A fully-insured health insurance contract is a contract purchased from an insurer where the insurer assumes the full risk of the individual. All other contracts are considered self-insured.

The American Medical Association (AMA) [2010] produces health insurance concentration figures for MSAs across the United States using HealthLeaders-InterStudy©. In general, we follow many of the AMA guidelines for calculating concentration measures using HealthLeaders-InterStudy© data. Specifically, we exclude PPO rental networks (e.g. Beech Street Corporation). These companies provide administrative services only and/or contract with health insurance carriers, which may cause double counting for those enrollees that are enrolled in another insurance plan that also contracts with a PPO rental network. We also exclude markets where HealthLeaders-InterStudy© data do not capture a plausible fraction of the insured population. Specifically, we calculate the ratio of total enrollment to the total eligible enrollment (i.e. population-uninsured-(Medicare+Medicaid-Dual)) estimated fraction of total possible enrollment in the market. Similar to the AMA, we exclude those MSAs where the ratio is less than 30 percent. Unlike the AMA concentration measures that only includes HMO and PPO enrollment we also include POS enrollment.

[43] As an alternative to the total enrollment, we also constructed an HHI based solely on the fully-insured insurance share information, and we obtain similar results.

information for U.S. counties: information on demographics, health facilities, health professionals, measures of resource scarcity, health status, and economic activity. The data are gathered from various sources, often on an annual basis.[44] The variables constructed from these data that are used in our analysis include median household income, education, population, population over the age of 65, hospital facility characteristics and a number of additional variables.[45]

4 Estimation of Period 1: Effects of Market Power on Service Price and Benefits

The goal of this study is to estimate if, and to what extent, physician market power dictates medical-care expenditures. As discussed in the previous section, one can categorize the determination of medical-care expenditures into three distinct periods of decision making. In Period 0, entry and exit of physician and insurance firms takes place. In Period 1, fees are negotiated and benefits are chosen, and in Period 2, a service utilization decision is made. In this section, we estimate the determinants of these Period 1 decision variables—the negotiated service price, P_n, and the benefit schedule, α_n. In estimating the determinants of the service price, we pay particular attention to the degree of bargaining leverage of physicians relative to insurance carriers. In estimating the determinants of the benefits schedule, we examine the impact of insurance carrier concentration. In the subsequent section, we will estimate the determinants of service utilization.

4.1 Determinants of Service Price

4.1.1 Specification

The following estimation routine quantifies the impact of the relative physician-insurance carrier bargaining leverage on the logarithm of service price P_n:

$$\begin{aligned} \ln(P_n) &= \beta_1 \ln(FTHHI_{phys}) + \beta_2 \ln(HHI_{ins}) + \delta' COST \\ &\quad + \kappa' QUAL + \lambda' PAT + \zeta_{at} + \zeta_d + \varepsilon_n. \end{aligned} \tag{12}$$

[44] Some of the sources included Census, the American Hospital Association database, American Medical Association database.

[45] Some of the additional variables include rental value of property, population over the age of 65 and share of hospitals that are university facilities.

where each episode, n, is uniquely associated with a patient k, a disease-stage-of-illness d, a physician p, an MSA m in a county c, and state a in time t. This specification essentially splits our measure of bargaining power leverage, Z, into its two components, $FTHHI_{phys}$ and HHI_{ins}.[46] As discussed above, the FTHHI of physicians is constructed at the physician-specialty level (see the Appendix), while the HHI of insurance firms is constructed at the MSA geographic level.

We include state-time fixed effects, ζ_{at}, as well as disease-stage-of-illness fixed-effects,[47] ζ_d, defined by the MEG. Specifically, the MEG algorithm classifies an episode of care into five major stages of illness and is meant to indicate the severity of a particular episode compared with other episodes in that disease group.[48] See Table 1 for summary statistics on this measure. To decrease computational burden, we include only the 100 most common disease groups for each specialty, which represents over 90 percent of the total samples.[49] We control for demographic attributes of the patient with the vector PAT, which includes a polynomial in the patient's age, a dummy variable indicating the patient's gender, as well as a polynomial in the number of co-morbidities of the patient. This latter variable is meant to control for those patients with multiple diseases, who may be sicker or harder to treat than patients with only a single disease. The patient-specific variables also include the patient's type of insurance carrier (for example, HMO, PPO, etc), whether the patient works for a larger employer[50], the logarithm of the median income of the patient's county, as well as the logarithm of the fraction of college educated individuals in the patient's county. We also include covariates that control for the physician's cost, $COST$, as well as the physician's quality, $QUAL$. The latter vector includes the percentage of hospitals in the physician's county that are affiliated with a university as well as a weighted average of the patient's county-level median income.[51] The former vector includes the logarithms of

[46] It would be equivalent to using Z as a covariate if we were to constrain $\beta_1 = -\beta_2$.

[47] For example, stage 3 acute myocardial infarction.

[48] Specifically, MEG assigns a severity score to each patient episode based on the "Disease Staging" disease progression model and does not depend on the utilization of care. Stage 0 represents a history or suspicion of a condition, exposure to a disease, or well visits. Stage 1 represents conditions with no complications or problems with minimal severity. Stage 2 represents problems limited to a single organ or system, significantly increased risk of complication than Stage 1. Stage 3 represents multiple site involvement, generalized systemic involvement, or poor prognosis and Stage 4 represents death.

[49] No results changed when all diseases were included on a 30 percent subsample of the data.

[50] This is actually based on an indicator of whether the data source for the claims information is from an employer (which is typically a large employer) or from a health insurance firm.

[51] The basic idea is that the higher quality doctors may attract the more wealthy patients. The weighted average is the average income of patients that see a particular doctor (based on county-level income data). This income variable is an average across patients seeing a particular doctor, which is distinct from the

the median rent, median home price, median income and average health care facility wage in the physician's county. These cost variables were chosen as controls because fees are usually bargained as a percentage of Medicare prices, which vary by cost-of-living variables such as rent, wages, and house prices. See Appendix B for details on the construction of these variables as well as specifications with additional controls for firm scale and physician supply.[52]

4.1.2 Correcting for Unobserved Bargaining Chips

The reason for including the controls described above is that they account for the many factors used in physician-insurance carrier bargaining. For instance, we include cost-of-living variables because physicians and insurers often bargain off of Medicare's relative value unit system. It is important to control for these cost-of-living factors because they may also affect the location choices of physicians and insurance firms. Estimation bias can still arise, however, if important bargaining chips exist that are unobserved to the econometrician and also affect competition. One example is unobserved physician quality. Higher quality physicians and facilities may result in larger negotiated fees, drawing more physicians into an area and dissuading insurance carriers from entering a market. More generally, potential bias may arise if physician firms are more likely and insurance carriers are less likely to enter markets where fees are high due to factors unobserved to the econometrician. This would result in upward bias on the insurance HHI coefficient and downward bias on the physician HHI coefficient. We have attempted to address this issue in two ways.

First, we constructed the $FTHHI$ using predicted market shares based on a fixed travel time as opposed to actual market shares. This approach is analogous to that taken by Kessler and McClellan [2000] who construct an HHI for hospitals using predicted market shares based on travel distances between patients and hospitals. Similarly, the predicted shares used in the $FTHHI$ measure are based solely on the number of physicians in the firm as well as the patient's travel time to the firm. As mentioned previously, Kessler and McClellan argue that this removes any endogeneity that may be attributable to higher quality providers attracting more patients.

patient income variable that enters as a demographic variable.

[52]We chose not to include these variables in our main specification because the are likely endogenous variables. Gaynor and Haas-Wilson [1999] note that "the extant literature on physician groups suggests that scale economies for such practices are also exhausted at relatively small sizes-three to five physicians (Pope and Burge [1996])."

Second, we implement two distinct instrumental variables strategies that are taken from the industrial organization literature—one taken from Berry and Waldfogel [2001] and Davis [2006a], and the other from Baker and Corts [1996]. The key to a good instrumental variable estimation strategy is to identify competitive variation solely attributable to the long-run entry and exit patterns of physicians and insurers—modeled as "Period 0" in our theoretical framework. Specifically, instruments that relate to either the fixed cost of entering the market or the market size are most appropriate, since they affect the entry decisions of insurers and physicians, but do not affect their pricing decision in Period 1, except through the impact on the number of rivals in the market. Berry and Waldfogel [2001] use population terms as instruments for radio broadcasting concentration, while Davis [2006a] uses population counts as instruments for his market structure variables in the movie-theater industry. Baker and Corts [1996] use firm size distribution variables to instrument for HMO market penetration.[53] We use two different instrument sets that are based on this literature. Both instrument sets take into account that for clean identification, we need instruments that are correlated with insurance entry and physician entry *uniquely*. Note that applying instrumental variables is especially important for identifying the impact of the insurance concentration measure, HHI_{ins}, since it is based on actual, rather than predicted, market shares.

Our first instrument set, which we label "population-distribution instruments," includes populations of certain age groups as well as different employment statuses. A greater number of senior citizens should encourage more entry from physicians relative to insurance carriers since these are generally sicker patients already covered by Medicare. We also include the population of employed individuals (the unemployment rate) for a similar reason. Insurance carriers are more likely to enter geographic areas with more employed individuals where the base of potential customers is higher. Physicians, however, are more indifferent to this factor because unemployed individuals are often covered by Medicaid or COBRA. In particular, we expect there to be a larger number of physician firms and insurance firms in more populated markets. Unlike prior work that uses population as instruments in a model of competition using aggregate data (for example, Davis [2006a] and Berry and Waldfogel [2001]), here we use detailed micro level data to control for

[53]Similar studies in the hospital literature have found it crucial to use instrumental variables to account for unobserved quality. As a recent example, Dranove et al. [2008] provides an instrumental variable strategy for estimating the effects of concentration on price in hospital markets.

Findings in the physician literature also suggest that price may be endogenous. Frank [1985] finds that psychiatrists respond to price; Schwartz et al. [1980], Newhouse et al. [1982], find that physicians locate in response to effective demand.

numerous factors at the level of the patient, so there is little reason to expect population to be correlated with physician quality. If, for some reason, higher quality physicians prefer to practice medicine in more populated areas, our estimates will be attenuated towards zero statistical significance. In this sense, our estimates will be conservative.[54]

Our second instrument set, which is labeled "firm-distribution instruments," consists of the number and size distribution of business establishments in the county. These variables are meant to capture the variation in competitive conditions of *all* industries in the market. We include the size distribution of the firms because insurance-carrier entry may be more affected by larger firms than smaller firms since larger firms are more likely to offer health insurance to their employees and presumably demand more insurance variety for its larger employee base.[55] The size distribution may also be related to the expected profitability from the type of insurance purchased by smaller firms, that tend to purchase full insurance, and larger firms, that tend to purchase self-insurance. Whereas the population-distribution instruments are designed to pick up cross-sectional variation in the exogenous characteristics of the market that determine entry, the firm-distribution instruments are designed to pick up the ex-post cross-sectional variation in the entry and exit of firms whose decisions are presumably exogenous to the physician and insurance market. The firm-distribution instruments may be related to both the size of the market, but also the fixed cost of entering a market that are common across firms of different industries.

In Appendix C we show estimates of the first-stage estimates of the instruments as well as results from a validity exercise. In all first-stage regressions, F statistics, testing the joint significance of the instruments, were large (ranging from 13 to 168).[56] As a validity check, we collected the second-stage residuals from (12) under the firm-distribution

[54]One may be concerned that population may be correlated with physicians per capita. However, we show in Appendix C.4.1 that results do not change when we include the number of physicians per capita as a control variable.

[55]In 2008, 96.5 percent of firms with more than 50 employees offered insurance, while only 43.2 percent of firms with fewer than 50 employees offered insurance (See the Medical Expenditure Panel Survey - Insurance Component Table I.A.2). Gruber and Lettau [2004] show similarly large differences in the offer rates of large and small firms, even after controlling for a multitude of other factors. In 2008 for those employees working for firms with fewer than 50 employees, only 26.8 percent are offered more than two insurance plans. In contrast, for employees working for firms with 50 or more employees, 73.7 percent are offered two or more health plans (See the Medical Expenditure Panel Survey - Insurance Component Table I.B.2.c).

[56]As a robustness exercise, we replaced the instrument set with a set that included the population over and under 65 for the MSA and for the county (four total instruments). No qualitative results changed, however standard errors grew a bit.

Table 2: Determinants of Service Price

	Cardiology			Orthopedics		
	OLS	IV		OLS	IV	
	(i)	(ii)	(iii)	(i)	(ii)	(iii)
$\ln(FTHHI_{phys})$	0.038***	0.105***	0.106***	0.033***	0.111***	0.171***
	(0.005)	(0.015)	(0.018)	(0.004)	(0.021)	(0.047)
$\ln(HHI_{ins})$	0.019**	-0.321***	-0.290**	0.025***	-0.240**	-0.556**
	(0.010)	(0.095)	(0.135)	(0.008)	(0.096)	(0.226)
Instruments	-	Pop. Dist.	Firm Dist.	-	Pop. Dist.	Firm Dist.
Observations	3668971	3664391	3668971	4133902	4129905	4133902

Notes: The dependent variable is the logarithm of service price, $\ln(P_n)$. All regressions include the controls specified in equation (12). Standard errors are in parentheses and are clustered by provider. One, two, and three asterisks indicate significance at the 10-percent, 5-percent, or 1-percent significance level, respectively.

instrument set, and ran a regression of these residuals on the population instrument set. If the residuals were highly correlated with the population-distribution instruments, this would suggest a potential bias. The residuals were not correlated ($R^2 \approx 0.001$) with the population-distribution instrument set (see Appendix C.3). Standard errors were all large, the R^2 of both samples was quite low, around 0.001, and F statistics testing their joint significance were also fairly low (1.8 and 2.7 for cardiology and orthopedics, respectively).

4.1.3 Results

We report results of specification (12) in Table 2. Standard errors are clustered by provider. This degree of clustering is meant to control for the fact that physicians bargain with an insurance carrier over an entire fee schedule.[57] In the cardiology as well as the orthopedic sample there is a positive and statistically significant effect of physician leverage on price per service. The OLS estimates indicate that a 10 percent increase in physician

[57] We also estimated a different specification of the episode price, P_n, regression where we used procedure price, p_j, as the dependent variable while including procedure fixed effects. This specification will be identical to specification (12) if physicians bargain with insurance carriers according to a discount on *all* procedures. That is, if $p_{jn} = \theta_n r_j \ \forall \ j$ for some $|\theta_n| < 1$, then it follows that $\ln(P_n) = \ln(\frac{p_{jn}}{r_j})$, which is equivalent to $\ln(\theta_n)$ as the dependent variable. No results changed using this specification indicating that, on average, physicians likely bargain over their entire fee schedule.

concentration will cause about a 0.3 percent increase in fees, but they also show that a 10 percent increase in insurance concentration causes a 0.2 percent increase in fees.

Using instrumental variables appears to remove the downward bias on $FTHHI_{phys}$ and the upward bias on HHI_{ins} that is likely attributable to unobserved bargaining chips. For the cardiology sample, estimates found using both instrument sets indicate that a 10 percent increase in physician concentration will result in about 1.1 percent higher fees for cardiologists, on average. For orthopedics, this effect lies in the 1.1 to 1.7 percent range, depending on the instrument set. Price effects from a change in the concentration of the insurance carrier are quite large, as a 10 percent increase in the insurance carrier's HHI will reduce prices by about 3 percent for cardiologists and 2 to 5 percent for orthopedists.[58]

4.2 Determinants of Benefits

As discussed in Section 2, a horizontal theory of competition among health insurers would suggest that markets with higher concentrations of insurance carriers likely offer plans with less generous benefits. To verify the effects of competition in this market, we run the following estimation routine, which quantifies the impact of insurance carrier concentration on our measure of the generosity of benefits, α_n:

$$\ln(\alpha_n) = \beta_3 \ln(\widehat{HHI_{ins}}) + \lambda' PAT + \zeta_{at} + \zeta_d + \varepsilon_n. \tag{13}$$

where we instrument for HHI_{ins} using either the population-distribution or firm-distribution instruments (12). Instrumenting may be important if there exist unobservable variables that are correlated with insurance carrier competition as well as benefit levels. For example, an unobservable that raises premiums or lowers benefits (higher α_n) may encourage more insurance entry, producing negative bias on β_3. Here we control for attributes of the patient with the vector PAT and disease-stage-of-illness fixed effects, which are included to control for any characteristic that may affect the patient's insurance carrier decision.[59] We also include state-time fixed effects, ζ_{at}.

[58] As an alternative to the OLS results, we also estimate the fee regression using county fixed effects and we obtain a similar coefficient on the physician $FTHHI_{phys}$ coefficient, although it is slightly lower. The county fixed effects will control for all factors unique to a provider in a county that are not captured by other variables. Although the county fixed effects make identification more difficult, we are still able to identify competitive effects from the fact that different providers compete in a different fashion for patients in neighboring counties.

[59] No results changed when we included the vector $COST$ and the vector $QUAL$.

Table 3: Determinants of Benefit Schedule

	Cardiology			Orthopedics		
	OLS	IV		OLS	IV	
	(i)	(ii)	(iii)	(i)	(ii)	(iii)
$\ln(HHI_{ins})$	0.069***	0.140	0.209*	0.043**	0.215***	0.236**
	(0.022)	(0.095)	(0.111)	(0.017)	(0.073)	(0.098)
Instruments	-	Pop. Dist.	Firm Dist.	-	Pop. Dist.	Firm Dist.
Observations	2977925	2974327	2977921	3824770	3821133	3824767

Notes: The dependent variable is the logarithm of the share of expenditures paid by the patient, $\ln(\alpha_n)$. All regressions include the controls listed equation (13). Standard errors are in parentheses and are clustered by MSA. One, two, and three asterisks indicate significance at the 10-percent, 5-percent, or 1-percent significance level, respectively.

Table 3 reports the estimate of β_3 for each sample under OLS and IV using the two distinct instrument sets. As expected, an increase in health-plan concentration is associated with a larger share of expenditures being paid out-of-pocket. In all specifications, there is a positive effect of insurance concentration on out-of-pocket shares. Under OLS, a ten percent rise in insurance carrier concentration is associated with a 0.7 percent increase in α_n for cardiology patients and 0.4 percent increase for orthopedic patients. Using instrumental variables raises the size of the coefficients, indicating there are likely unobservable variables that encourage insurance entry that also worsen benefits. Standard errors are relatively large but should be considered conservative since we are clustering at the MSA level in this specification.

The previous literature offers relatively little evidence of the effects of health insurance competition on consumer welfare. The findings in this section provide an important contribution to the literature by showing that insurers in more consolidated markets are able to reduce medical benefits to consumers, which is consistent with the recent work by Dunn [2010] and Dafny et al. [2011] who find that additional consolidation leads to higher premiums and lower benefits.

5 Estimation of Period 2: Service Utilization

Having estimated the determinants of the first-period variables—service prices and health-plan benefits—we are now in a position to estimate the determinants of the utilization of services. As delineated in the earlier part of this paper, the utilization of services is decided upon by the physician and the patient, given the first-period negotiated prices and chosen benefit schedule. A key to our identification strategy is that service price changes should affect utilization differently depending on the location of the patient's demand curve. Among those patients who are not price sensitive and have a high demand for services, a movement in price will likely represent a movement along the physician's supply curve. By contrast, for those patients who are price sensitive and have a lower demand for services, shifts in price will likely represent movements along the demand curve. For expositional purposes, we motivate our structural specification with a simple subsample analysis. We then move to a structural switching regression specification. The following subsections also discuss the empirical issues that arise when analyzing these relationships.

5.1 Subsample Analysis

To begin our analysis of service utilization, note that we can estimate the effect of a marginal price change on service utilization as:

$$\ln(Q_n) = \gamma \ln(P_n) + \delta' COST + \kappa' QUAL + \lambda' PAT + \zeta_{at} + \zeta_d + \varepsilon_n, \tag{14}$$

where γ is a coefficient representing a price elasticity of service utilization. As a first exercise we perform a simple subsample analysis using equation (14). Our subsample will be based on the intuition behind Figure (1), which implies that the econometrician can estimate the supply elasticity by assessing the effect of a change in price on utilization for those individuals with generous benefits, who should be less sensitive to movements in market price. Similarly, the demand elasticity may be estimated by measuring the effect on those individuals with less generous benefits, who should be more sensitive to movements in market price. Intuitively, one may think that the level of coverage of an individual may be a key determinant of whether the binding constraint is from the physician or the patient. Thus, we estimate specification (14) on two distinct subsamples or "regimes:" those observations with α_n above the median that are more likely to capture patient demand (regime-D) and those observations with α_n below the median that are likely to represent observations on the physician marginal cost curve (regime-S).

In this simple framework, an estimate of the demand elasticity, γ^D, can be found by estimating (14) on regime-D and an estimate of the supply elasticity, γ^S, can be found by estimating (14) on regime-S. Having identical covariates in each regime means that the estimate of the demand elasticity, γ^D, is identified solely from variation in the service price, P_n. That is, we are estimating a demand function that takes the form $D(\bar{\alpha}P_n)$, where α is assumed to be fixed. The benefit to this approach is that it removes a great deal of endogenous variation attributable to α_n in the estimation routine.

5.1.1 Correcting for Unobserved Selection, Cost, and Quality

There are numerous potential endogeneity biases in modeling both the physician's and the patient's response to price. The physician's response to the negotiated service price on utilization is clearly an endogenous relationship, since a higher negotiated price may simply reflect higher quality. This may lead to a positive or negative bias in the relationship between P_n and utilization in the price coefficient. Alternatively, a higher negotiated price may reflect higher unobserved physician costs, which would introduce a negative bias in this relationship. Another endogeneity issue arises with the measure of α_n, which is used to divide the sample into two subsamples. As α_n is constructed by dividing out-of-pocket payments by total payments, any nonlinear structure of benefits attributable to deductibles and maximum dollar expenditures means that α_n is inherently dependent on the underlying quantity of services provided.[60]

These potential biases are addressed using both the population-distribution and firm-distribution instruments. As shown in the previous sections, these instruments are related to the service price and benefit generosity through the determinants of physician and insurer entry in Period 0. In particular, the competitive effects estimates confirm that the instruments are strongly correlated to both the negotiated price and benefits through shifts in the competitive environment. More generally, the instruments may be considered "upstream instruments" since they affect Period 0 variables (market structure), which affect Period 1 variables (price and benefits), which subsequently affect utilization in Period 2.[61] In sum, these will be good instruments insofar as they are correlated with price and benefits solely due to the long-run competitive patterns of physicians and insurance

[60] For instance, episodes with a very large quantity of services may inherently have a low α_n, even though the actual benefit structure may be the same as a patient with a larger α_n but a more moderate degree of services provided.

[61] A similar argument regarding upstream instruments was used to motivate an identification strategy in Shapiro [2008].

carriers determined in Period 0. In our estimation routine, we instrument for $\ln(P_n)$ in both subsamples. Also, due to endogeneity issues of using, α_n, to split the sample, we use predicted, rather than actual α_n, in dividing the sample.[62]

We should emphasize that we are estimating *individual* demand (which we label "utilization") which measures the amount of services *conditional* upon being treated. This means that *aggregate* population does not enter the service utilization specification.[63] Since we are able to condition on detailed individual-specific information (for example age, sex, disease, and stage of illness), the population-distribution instruments should only be related to the utilization of services through its effect on price and benefits. Similar arguments have been made in Kennan [1989], Gaynor and Vogt [2003], and Dunn [2012].[64] Population instruments have also been applied to study demand in markets using more aggregate data (for example, Berry and Waldfogel [2001], Rysman [2004], Davis [2006b], and Romeo [2010]). As an important check on the population instruments, we also estimate the utilization regressions using the firm-size distribution instrument set, which is arguably much less related to illness severity in a market.[65]

5.1.2 Results

Estimates of γ^D (measured as the coefficient on $\ln(P_n)$ in the sample where $\hat{\alpha}_n > \bar{\hat{\alpha}}_n$) and γ^S (measured as the coefficient on $\ln(P_n)$ in the sample where $\hat{\alpha}_n < \bar{\hat{\alpha}}_n$) are shown in Table 4. We also show estimates of the full sample. In line with the theoretical framework, the effect of a change in service price on utilization is dependent on the generosity of benefits. However, standard errors are quite large and our estimates of the demand elasticity

[62] We divide the sample based on predicted measures, $\widehat{\alpha_n}$, the fitted values from $\ln(\alpha) = \beta' IV + \theta' COST + \kappa' QUAL + \lambda' PAT + \zeta_{at} + \zeta_d + \varepsilon_n$, where IV is the set of instruments. For robustness purposes, we also used predicted measures based on $\ln(\alpha) = \beta' IV + \varepsilon_n$. Results did not qualitatively change.

[63] Aggregate demand may be derived by multiplying the individual predicted demand by the number of individuals.

[64] To better understand how these instruments function, it is useful to note how they may fail. In particular, suppose that we did not condition on the individual specific information on age. In this case, the age distribution in the population may be related to the individual's unobserved illness severity, which may also be positively related to price, introducing a positive bias in the relationship between price and utilization.

[65] We should note that this instrumenting strategy differs from a common exclusion restriction in demand estimation that uses "cost shifters" (see Bresnahan [1989]). However, employing a "cost shifter" as an instrument will be potentially problematic in the switching framework that follows. At issue is that in estimating supply relationship, costs must be held fixed. Therefore, a cost shifter would clearly introduce a bias on the physician's price coefficient because it is directly related to unobserved costs.

Table 4: Subsample Analysis

	Cardiology		Orthopedics	
γ^D	0.501	0.179	-0.255	-0.112
	(0.318)	(0.292)	(0.165)	(0.134)
γ^S	1.112***	0.281	0.349*	0.356**
	(0.368)	(0.295)	(0.179)	(0.176)
Full Sample	0.875***	0.283	0.120	0.134
	(0.334)	((0.276)	(0.159)	(0.134)
Instruments	Pop. Dist.	Firm Dist.	Pop. Dist.	Firm Dist.

Notes: γ^D is estimated as the coefficient of $\ln(P_n)$ from specification (14) in the sample where $\hat{\alpha}_n > \bar{\hat{\alpha}}_n$. Similarly, γ^S is estimated as the coefficient of $\ln(P_n)$ from specification (14) in the sample where $\hat{\alpha}_n < \bar{\hat{\alpha}}_n$. All regressions include the controls specified in equation (14). Standard errors are in parentheses and are clustered by provider. One, two, and three asterisks indicate significance at the 10-percent, 5-percent, or 1-percent significance level, respectively.

are the correct sign (negative) only for the orthopedic sample.

The estimates of γ^D in the orthopedics sample are -0.26 using the population-distribution and -0.11 using the firm-distribution instruments, both of which are statistically insignificant from zero. The estimates of γ^S are 0.35 and 0.36, respectively. In the cardiology sample, the estimates of γ^D are positive and both statistically insignificant. In all cases, the estimate of γ^S is larger than γ^D, which provides evidence that cost-sharing with the patient does in fact dampen the effect of the physician's positive supply elasticity.

5.2 Switching Regression

According to our stylized example in Figure 1, whether the patient's demand curve is binding depends on the level of benefits α_n as well as the service price P_n, but realistically it may also depend on other covariates that determine relative service demand such as plan type and disease. With this in mind, an obvious drawback of performing the subsample analysis is that it imposes *exactly* which observations are constrained by the demand curve and supply curve. A more flexible approach would be to estimate which observations are constrained. We address this issue by employing an E-M algorithm proposed by Kiefer (1980). In this empirical framework, the probability that the patient-episode-physician

Table 5: Switching Regression

	Cardiology		Orthopedics	
γ^D	-0.049	-0.283**	-0.429***	-0.321***
	(0.099)	(0.092)	(0.109)	(0.101)
γ^S	1.256***	0.567*	0.267*	0.336***
	(0.328)	(0.274)	(0.141)	(0.120)
$Pr(\omega_n = 1)$	0.340	0.302	0.610	0.607
Instruments	Pop. Dist.	Firm Dist.	Pop. Dist.	Firm Dist.
Log Likelihood	-6.26e6	-6.15e6	-6.81e6	-6.83e6
Observations	3664391	3669333	4131763	4131592

Notes: Standard errors are in parentheses and are clustered by provider. One, two, and three asterisks indicate significance at the 10-percent, 5-percent, or 1-percent significance level, respectively.

triple (n) belongs to a particular regime is, $Pr(\omega_n = 1|X)$, estimated along with equation (14) on both the regimes. Here, $\omega_1 = 1$ indicates we are in regime-D where the patient's constraint is binding, and X indicates the vector of covariates in equation (14).

Since we have identical covariates for each regime, identification of distinct parameters stems from the assumption that the initial starting distribution of ω_n provides a "close-enough" approximation to the distinct data-generating distributions of the two regimes. We chose starting values for ω_n in an analogous fashion to how we divided the sample in the subsample analysis. Specifically, we chose starting values of $\omega_n = 1$ if $\hat{\alpha}_n$ lies above the median, and $\omega_n = 0$ if $\hat{\alpha}_n$ lies below the median. In this manner, the switching framework can be viewed as a $refinement$ of the subsample analysis above, whereby the E-M algorithm searches for the mean and variance of the two actual data-generating distributions. Once the algorithm begins, $Pr(\omega_n = 1|X)$ is free to follow any particular pattern.

Estimates of the switching framework are shown in Table 5. The results on the orthopedics sample look similar to the subsample analysis, however, standard errors are considerably smaller. The demand elasticity estimates on cardiology are now the correct sign, albeit, statistically significant only using the firm-distribution instruments. For orthopedics, the estimate of the demand elasticity is -0.427 for the population-distribution instruments and -0.321 for the firm-distribution instruments. The supply elasticity estimates are 0.275 and 0.336, respectively. Estimates on the cardiology sample are somewhat less stable than those for orthopedics, however, they show qualitatively similar results. The estimate of the demand elasticity is -0.049 using the population-distribution instruments and -0.283 using the firm-distribution instruments. Both estimates imply that patients are price sensitive, but relatively inelastic. The supply elasticity estimates are considerably larger for cardiology than for orthopedics: 1.256 and 0.556, for the population-distribution and firm-distribution instruments, respectively. Overall, the estimates indicate that elasticities are relatively small on both the patient and physician sides of the market. These estimates for orthopedics are close in magnitude to those measured using randomized data from the RAND health insurance experiment (Manning et al. [1987] and Keeler and Rolph [1988]) who find elasticities in the -0.1 to -0.2 range.[66]

In the orthopedics sample, the probability that the demand-curve is binding is around 0.6 indicating that a slight majority of orthopedic patients are price sensitive. In the cardiology sample, the probability that the demand-curve is binding is around 0.3 indicating that most cardiology patients are not price sensitive. Thus, our estimates indicate that orthopedic patients are not only more price sensitive, but a larger fraction of orthopedic patients are price sensitive relative to cardiology patients. This result makes intuitive sense if one takes into account the fact that orthopedists generally do not treat life-threatening illnesses.[67]

In Appendix D we report estimates from a switching regression analysis that measures the patient's response to out-of-pocket price—as opposed to the patient's response to service price, discussed in this section. The estimates show out-of-pocket demand elasticities that are somewhat larger (in absolute value), however, the estimates also show smaller proportions of price sensitive individuals. These subtle differences may be a moot point in terms of the focus of this study because we found that both sets of estimates imply similar outcomes in terms of the effect of changes in the $FTHHI$ on service utilization. We discuss this topic in the next section.

[66]To calculate a service price elasticity for the full population of patients, one must multiply the patient elasticity coefficient by the expected probability that the patient's constraint is binding, $Pr(\omega_n = 1)$. For example, the patient elasticities for orthopedics using the two instrument sets are -0.262 (=0.610·-0.429) and -0.195 (=-0.321·0.607). These figures are more comparable to those in the RAND experiment.

[67]To additionally assess ω_n, we ran a regression of the predicted regime as the explanatory variables, X, as well as α_n and P_n. Consistent with the theoretical model, the estimated coefficients on both of these variables were positive and statistically significant in all four specifications.

6 Market Power and Service Provision

As described by the theoretical model, an increase in physician bargaining leverage can translate into either a higher or lower service utilization rate through its effect on service price. Our estimates from Section 4 imply that an increase in physician concentration raises service prices, while the estimates from Section 5 imply that an increase in service price will either raise or lower utilization.

The effect of market power on service provision can be calculated from the estimates in the previous two sections in a very simple way. Note that from the equation (12) the marginal effect of a change in $FTHHI_{phys}$ on $\ln(P)$ is β_1. Let γ^D represent the estimate of the demand elasticity—which is binding for those observations with $\omega_n = 1$—and let γ^S represent the estimate of the supply elasticity—which is binding for those observations with $\omega_n = 0$. It follows that the marginal effect of bargaining leverage on service utilization can be calculated as a weighted average:

$$\frac{\partial \ln(Q)}{\partial \ln(FTHHI_{phys})} = Pr(\omega_n = 1|X) \cdot \beta_1 \cdot \gamma^D + [1 - Pr(\omega_n = 1|X)] \cdot \beta_1 \cdot \gamma^S. \tag{15}$$

This equation shows that an increase in physician concentration will cause a more negative (positive) effect on utilization, the larger (smaller) is the absolute value of the service demand elasticity (γ^D) and the smaller (larger) is the service supply elasticity (γ^S). The effect will also become more negative (positive) the greater (fewer) the number of individuals in the market where the service demand curve is binding (that is, the larger (smaller) is $Pr(\omega_n = 1|X)$).

Plugging in the estimates from Table 5, this translates into a marginal effect of $FTHHI_{phys}$ on $\ln(Q_n)$ that is 0.032 (firm-distribution instruments) and 0.085 (population-distribution instruments) for cardiology and a -0.011 (firm-distribution instruments) and -0.017 (population-distribution instruments) for orthopedics. The positive effect for cardiology is attributable to the relatively larger supply elasticity as well as the lower proportion of price sensitive individuals in the cardiology sample. In the orthopedic sample, there is a relatively larger demand response. The effect of HHI_{ins} on utilization can be calculated by substituting β_2 for β_1:

$$\frac{\partial \ln(Q)}{\partial \ln(HHI_{ins})} = Pr(\omega_n = 1|X) \cdot \beta_2 \cdot \gamma^D + [1 - Pr(\omega_n = 1|X)] \cdot \beta_2 \cdot \gamma^S, \tag{16}$$

Plugging in estimates from Table 5, this translates into a marginal effect of -0.089 (firm-distribution instruments) and -0.261 (population-distribution instruments) for cardiology

Table 6: Market Structure and Service Utilization

	Cardiology			Orthopedics		
	OLS	IV		OLS	IV	
	(i)	(ii)	(iii)	(i)	(ii)	(iii)
$\ln(FTHHI_{phys})$	0.001	0.095***	0.072	-0.018***	0.023	0.040
	(0.015)	(0.021)	(0.046)	(0.005)	(0.025)	(0.042)
$\ln(HHI_{ins})$	-0.079***	-0.441***	-0.378	-0.006	-0.172	-0.254
	(0.027)	(0.114)	(0.307)	(0.012)	(0.106)	(0.213)
Instruments	-	Pop. Dist.	Firm Dist.	-	Pop. Dist.	Firm Dist.
Observations	3668971	3664391	3668971	4133902	4129905	4133902

Notes: The dependent variable is the logarithm of the utilization of services, $\ln(Q_n)$. All regressions include the controls specified in equation (17). Standard errors are clustered by provider. One, two, and three asterisks indicate significance at the 10-percent, 5-percent, or 1-percent significance level, respectively.

and 0.034 (firm-distribution instruments) and 0.036 (population-distribution instruments) for orthopedics.[68]

Analogous to estimating the effect of concentration measures on service price, as done in Section 4, we can directly estimate the marginal effect of bargaining leverage on service utilization. Specifically, we estimate a reduced-form specification:

$$\begin{aligned} \ln(Q_n) &= \phi_1 \ln(\widehat{FTHHI}_{phys}) + \phi_2 \ln(\widehat{HHI}_{ins}) + \delta' COST \\ &\quad + \kappa' QUAL + \lambda' PAT + \zeta_{at} + \zeta_d + \varepsilon_n. \end{aligned} \tag{17}$$

where ϕ_1 is an approximation of $\frac{\partial \ln(Q)}{\partial \ln(FTHHI_{phys})}$ and ϕ_2 is an approximation of $\frac{\partial \ln(Q)}{\partial \ln(HHI_{ins})}$. As in our previous specification, we include disease-stage-of-illness fixed effects, ζ_d, state-time fixed effects, ζ_{at}, as well as controls for the physician's cost, $COST$, quality, $QUAL$, and the patient's demographic factors, PAT.

[68]The predicted effect of insurance concentration on utilization is different if one uses the switching specification that includes P^{pock} as shown in the appendix. Specifically, since $\ln(P^{pock}) = \ln(P) + \ln(\alpha)$, it follows that $\frac{\partial \ln(Q)}{\partial \ln(HHI_{ins})} = Pr(\omega_n = 1|X) \cdot \beta_2 \cdot [\gamma^D + \beta_3] + [1 - Pr(\omega_n = 1|X)] \cdot \beta_2 \cdot \gamma^S$ where $\beta_3 = \frac{\partial \ln(\alpha)}{\partial \ln(HHI_{ins})}$. The estimates on the insurance carrier side, β_3, indicate that an increase in health-plan concentration lowers benefits, as well as price. Thus, our estimates imply that insurance carriers can lower service utilization through both removing price incentives for physicians (via lowering P_n), as well as inducing patients to become more price sensitive (via raising α_n). Allowing for this effect would tend to make $\frac{\partial \ln(Q)}{\partial \ln(HHI_{ins})}$ more negative than the amount predicted by the equation in the text.

Estimates are depicted in Table 6. In general, the magnitude of the ϕ_1 estimate is consistent with our expectations based on the estimates in the previous two sections; while the ϕ_2 estimates tend to be more negative than the predicted amount, they also have relatively large standard errors.[69] The estimates of ϕ_1 for cardiology are 0.095 (statistically significant) using the population-instruments and 0.072 (statistically insignificant) using the firm-distribution instruments. Both estimates for orthopedics are small and positive, but not statistically significant. For cardiologists, the positive effect on utilizations means that the effects of physician bargaining power on medical-care expenditures become magnified relative to the scenario where bargaining power translates only into price effects. This can be seen more clearly by re-examining equation (11), which shows that, by construction, the sum of the coefficients on price and utilization will equal the coefficient on total episode expenditures. This implies that a 10 percent rise in $FTHHI_{phys}$ is associated with roughly a 2 percent increase in expenditures for cardiologists—roughly half of which is due to the effect on utilization.[70] To put these numbers in better perspective, all else equal, a cardiologist with the 90th-percentile $FTHHI_{phys}$ will have about 56 percent higher expenditures per episode on average than a cardiologist with the 10th-percentile $FTHHI_{phys}$. Splitting this number between the price and utilization component, the 90th-percentile cardiologist will charge 26 percent higher prices and perform 24 percent more services.[71]

More generally, absent the physician's influence on utilization (i.e. $\gamma^S = 0$), the unambiguous prediction of our model would be for an increase in market power to lead to a reduction in utilization as patients respond to higher prices. However, incorporating the physician incentives and their influence on utilization, we find that higher margins from consolidation actually lead to either no change or, in some cases, an expansion of services. This finding is quite distinct from most other markets where higher prices from consolidation tend to lead to a reduction in purchases.

[69]The more negative amount for $\frac{\partial \ln(Q)}{\partial \ln(HHI_{ins})}$ reported in Table 6 may be attributed to the fact that the above prediction based on the previous two sections ignores the effect of insurer concentration on benefits and subsequent utilization levels. Accounting for this additional effect would make the predicted effect more negative and consistent with the amount reported in Table 6.

[70]That is, $0.20 \approx 0.105 + 0.095$

[71]For example, $0.26 = \frac{exp(.105{*}\ln(.241)) - exp(.105{*}\ln(.026))}{exp(.105{*}\ln(.026))}$ where 0.241 and 0.026 are the 90th and 10th percentile $FTHHI_{phys}$, respectively.

7 Conclusion

The effects of physician bargaining power are important given the observed consolidation of physicians over the past few decades, and the potential increase in consolidation due to health care reform. This paper studies the role of physician bargaining leverage in determining service prices and service utilization—the two components of physician medical-care expenditures. Our estimates suggest that those physicians with greater market power relative to insurance carriers are able to receive higher service payments. Unlike typical markets, these higher payments do not correspond with lower utilization and may in fact increase utilization. We attribute this result to a low proportion of price sensitive patients as well as the presence of an upward sloping supply curve. Market power of insurance carriers also plays an important role. We provide evidence that insurance carriers with greater market power are able to negotiate lower service prices and are also able to reduce the generosity of physician benefits. These results have broad implications for antitrust policy, the structure of payment schedules to physicians, and the benefit design of insurance products.

These findings may explain a portion of the large geographic variation in overall medical expenditures documented in Dunn, Shapiro, and Liebman [2011]. In particular, this study shows how bargaining power of physicians and insurers may affect both the service prices and utilization of services. However, Dunn, Shapiro, and Liebman [2011] also document significant variation in many other health services (for example, hospital outpatient, hospital inpatient, and pharmacy services). Given the central role of physicians as the agents selecting health care services for patients, it is possible that the incentives of physicians may impact ancillary health care services used in the treatment of a disease (e.g. inpatient facility payments to a hospital). Expenditures may increase with physician utilization due to complementarities with other services, or physician services may be an alternative substitute for other treatments. Further research would entail analyzing how physician market power manifests itself into different mixes of services (for example, pharmacy services, inpatient services, outpatient services) being administered to the patient.

In sum, we find that the overall effects of physician market power on medical-care spending are large. However, we do not have information on the health outcome of the patient. Thus, it is not entirely clear whether those patients being treated by physicians with larger market power are receiving higher quality treatment and/or experiencing better health outcomes. Another promising area for future research would entail measuring how physician market power translates into physician quality and health outcomes.

References

[1] Aizcorbe, A. and Nestoriak, N. 2011. "Changing Mix of Medical Care Services: Stylized Facts and Implications for Price Indexes", *Journal of Health Economics.*

[2] Baker, L.C. and Corts, K. 1996. "HMO Penetration and the Cost of Health Care: Market Discipline or Market Segmentation?" *The American Economic Review*, Vol. 86, No. 2, Papers and Proceedings. pp. 389-394.

[3] Berry, S.T. and J. Waldfogel. 1999. "Free entry and Social Inefficiency in Radio Broadcasting." *The RAND Journal of Economics*, Vol. 30, No. 3, pp. 397-420.

[4] Berry, S.T. and J. Waldfogel. 2001. "Do Mergers Increase Product Variety? Evidence from Radio Broadcasting." *The Quarterly Journal of Economics*, Vol. 116, No.3, pp. 1009-1025.

[5] Bresnahan, T. and P. Reiss. 1991. "Entry and Competition in Concentrated Markets." *Journal of Political Economy.* Vol. 99. No. 5. pp 977-1009.

[6] Capps, C., D. Dranove and Satterthwaite, M. 2003. "Competition and Market Power in Option Demand Markets." *The RAND Journal of Economics.* 34(4): 737-763

[7] Cutler, D. 1995. "The Incidence of Adverse Medical Outcomes Under Prospective Payment." *Econometrica* 63(1), pp. 29-50.

[8] Cutler, D. and Ly, D. 2011. "The (Paper)Work of Medicine: Understanding International Medical Costs." *Journal of Economic Perspectives* 25(2) pp. 3-25.

[9] Chernew, M., Lindsay S., Chandra,A. Gibson, T. and Newhouse,J. 2010. "Geographic Correlation Between Large-Firm Commercial Spending and Medicare Spending", *American Journal of Managed Care*, 16(2) pgs 131-138.

[10] Chernew, M., Hirth, R., Cutler D. 2009 "Increased spending on health care: long-term implications for the nation." *Health Affairs* vol. 28 no. 5 1253-1255

[11] Dafny, L., M. Duggan, and S. Ramanarayanan, 2011. "Paying a Premium on Your Premium? Consolidation in the U.S. Health Insurance Industry." *American Economic Review.*

[12] Davis, P., 2005. "The Effect of Local Competition on Admission Prices in the U.S. Motion Picture Exhibition Market" *Journal of Law and Economics*, 48 pgs 677-708.

[13] Davis, P. 2006a. "Measuring the Business Stealing, Cannibalization and Market Expansion Effects of Entry in the U.S. Motion Picutre Exhibition Market." *Journal of Industrial Economics.*

[14] Davis, P. 2006b. "Spatial Competition in Retail Markets: Movie Thaters." *The RAND Journal of Economics*, Vol. 37. No. 4, pp. 964-982.

[15] Decker, S. 2009. "Changes in Medicaid Physician Fees and Patterns of Ambulatory Care." *Inquiry* Vol. 46, No. 3, pp. 291-304.

[16] Dranove, D., 1988. "Demand Inducement and the Physician/Patient Relationship." *Economic Inquiry.* Vol. 26 pgs. 251-298.

[17] Dranove, D. and R. Ludwick, 1999. "Competition and pricing by nonprofit hospitals: a reassessment of Lynk's analysis" *Journal of Health Economics* , 18 (1) , pp. 87–98

[18] Dranove, D., Shanley, M., and White, W. 1993. "Price and Concentration in Hospital Markets: The Switch from Patient Driven to Payer-Driven Competition." *Journal of Law and Economics*, Vol. 36, pp. 179-204.

[19] Dranove, D., R. Lindrooth, W. White, and J. Zwanziger, 2008. "Is the impact of managed care on hospital prices decreasing?", *Journal of Health Economics*, Vol. 27, pp. 362-376.

[20] Dranove, D., A. Gron, and M. Mazzeo, 2003. "Differentiation and Competition in HMO Markets." *Journal of Industrial Economics.*

[21] Duggan, M. 2002. "Hospital Market Structure and the Behavior of Not-for-Profit Hospitals", *RAND Journal of Economics*, 33:3, 433-446.

[22] Dunn, A., A.H. Shapiro, and E. Liebman, 2011. "Geographic Variation in Commercial Medical Care Expenditures: A Decomposition Between Price and Utilization." Working Paper.

[23] Dunn, A., 2011. "The Effect of Health Insurance Competition when Private Insurers Compete with a Public Option." Working Paper.

[24] Dunn, A., 2012. "Drug Innovations and Welfare Measures Computed from Market Demand: The Case of Anti-cholesterol Drugs" Forthcoming *American Economic Journal: Applied Economics*

[25] Evans, R., 1974. “Supplier-induced Demand” in: M. Perlman, ed., The Economics of Health and Medical Care (Macmillan, London) pgs 162-173.

[26] Federal Trade Commision (FTC) and Department of Justice (DOJ), 2004. Improving Health Care: A Dose of Competition. Report.

[27] Fuchs, V.R., 1978. “The Supply of Surgeons and the Demand for Operations”, *The Journal of Human Resources* pgs 35-56.

[28] Liebhaber A, Grossman JM., 2008. “Physicians Moving to Mid-Sized, Single-Specialty Practices.” Health Systems Change Tracking Report Report No. 18. www.hschange.org/CONTENT/941.

[29] Gaynor, M., and Haas-Wilson, D. 1999. “Change, consolidation, and competition in health caremarkets.” *Journal of Economic Perspectives*, 13(1), 141–164.

[30] Grant, Da. 2009. “Physician Financial Incentives and Cesarean Delivery: New Conclusion from the Healthcare Cost and Utilization Project.” *Journal of Health Economics*. Volume 28, Issue 1, Pages 244-250.

[31] Gruber, J., and M. Lettau. 2004. “How Elastic Is the Firm’s Demand for Health Insurance.” *Journal of Health Economics*. Volume 18, Issue 4, August 1999, Pages 473-490.

[32] Gruber, J., Kim, J., and Mayzlin, D. 1999. “Physician Fees and Procedure Intensity: The Case of Cesarean Delivery.” *Journal of Health Economics*. Volume 18, Issue 4, August 1999, Pages 473-490.

[33] Hemenway, D., Killen, A., Cashman, S. Parks, C., Bicknell, W. 1990. “Physicians’ Responses to Financial Incentives. Evidence from a For-Profit Ambulatory Care Center” *New England Journal of Medicine* Apr 12;322(15):1059-63.

[34] Haas-Wilson, D. 2003. *Managed care and monopoly power: The antitrust challenge*. Cambridge: Harvard University Press.

[35] Ho, K. 2009. “Insurer-Provider Networks in the Medical Care Market” *American Economic Review*, 99(1): 393-430.

[36] Horn, H. and A. Wolinsky. 1988. “Bilateral Monopolies and Incentives for Merger.” *The RAND Journal of Economics*, Vol. 19, No. 3, pp. 408-419

[37] Kiefer, N. 1980. "A Note on Switching Regressions and Logistic Discrimination." *Econometrica* Vol. 48. No.4. pp 1065-1069.

[38] Keeler, E.B., Melnick, G. and Zwanziger, J. 1999. "The Changing Effects of Competition on Non-profit and For-Profit Hospital Pricing Behavior." *Journal of Health Economics*, Vol. 18, pp. 69-86.

[39] Keeler, E. and Rolph, J. 1988. "The Demand for Episodes of Treatment in the Health Insurance Experiement." *Journal of Health Econoimcs*,

[40] Kessler, D., and M. McClellan. 2000. "Is Hospital Competition Socially Wasteful?", *The Quarterly Journal of Economics*, Vol. 115. No. 2. pp. 577-615

[41] Lynk, W. 1995. "Nonprofit Hospital Mergers and the Exercise fo Market Power." *Journal of Law and Economics*

[42] McGuire, T.G. and M.V. Pauly, 1991. "Physician Response to Fee Changes with Multiple Players.", *Journal of Health Economics* pp. 385-410.

[43] Noether J. 1988. "Competition Among Hospitals." *Journal of Health Economics*, Vol. 7 , pp. 259-284.

[44] Pauly, M.V., 1980. *Doctors and Their Workshops: Economic Models of Physician Behavior*. University of Chicago Press, Chicago.

[45] Pope, G. C., Burge, R. T., 1996. "Economies of scale in physician practice." *Medical Care Research and Review*, 53(4), 417-440.

[46] Rebitzer, J. and M. Votruba. 2011. "Organizational Economics and Physician Practices." Working Paper 17535, NBER Working Paper Series.

[47] Romeo, C. 2010. "Filling Out the Instrument Set in Mixed Logit Demand Systems for Aggregate Data." Workin Paper.

[48] Rysman, M. 2004. "Competition Between Networks: A Study of the Market for Yellow Pages." *Review of Economic Studies*, 71, 483-512

[49] Shapiro, A.H. 2008. "Estimating the New Keynesian Phillips Curve: A Vertical Production Chain Approach." *Journal of Money, Credit, and Banking*

[50] Sirovich B, Gallagher PM, Wennberg DE, Fisher ES. 2008. "Discretionary decision making by primary care physicians and the cost of U.S. Health care." *Health Affairs* 2008 May-Jun; 27(3):813-23.

[51] Smart D.R. 2006. *Medical group practices in the US*, 2006 edition. Chicago, IL: American Medical Association.

[52] Sorensen, A. 2003. "Insurer-hospital bargaining: negotiated discounts in post-deregulation Connecticut." *The Journal of Industrial Economics*, 51: 469–490.

[53] Staten, M. Umbeck, J. and Dunkelberg, W., 1987. "Market Share and the Illusion of Power: Can Blue Cross Force Hospitals to Discount?" *Journal of Health Economics* 6(1), 43-58.

[54] Staten, M. Umbeck, J. and Dunkelberg, W., 1988. "Market Share/Market Power Revisited: A New Test for an Old Theory," *Journal of Health Economics* 7(1), 73-83.

[55] Smart DR. Medical group practices in the US, 2006 edition. Chicago, IL: American Medical Association; 2006

[56] Town, R. and Vistnes, G. 2001. "Hospital Competition in HMO Networks," *Journal of Health Economics*, Vol. 20., pp. 733-753.

Appendix

A Construction of Fixed-Travel-Time HHI

We construct fixed-travel-time concentration measures in the following fashion. For each geographic location we define a latitude and longitude location as a vector $x = \{lat, long\}$. Using Google's Maps software we can measure a maximum radial distance based on amount driving time $\bar{k}$. For instance, for any location x we can calculate a radius of $\bar{k} = 80$ minutes of driving time. To do so, for each county, c, we drew a random coordinate and then calculated the average speed, $speed_c$, one could travel 0.1 degrees north, south, east, and west latitude. We use the Stata package, "traveltime," written by Ozimek and Miles. This allows us to define a maximum radial distance for any latitude and longitude coordinate in county c as $\bar{k} * speed_c$.

For each location in the SK&A data, we attach weights to each physician group in the surrounding area. These weights can be thought of as probabilities of whether a patient located at x_0 would be willing to travel to a physician located at x_i. For a patient located at x_0, we define their driving time to a physician located at x_i as k_{x_i}. It follows that a patient who lives at location x_0 resides k_{x_i} minutes away from the physician located at x_i. We then create a weight variable which represents the probability that a patient located at x_0 would consider traveling to the physician located at x_i. We do this in the most tractable manner possible by assuming that patients' idiosyncratic taste shocks lie on the uniform distribution and that k_{x_i} is directly proportional to travel costs. Specifically, a patient will choose a physician located at x_i instead of a physician located at x_0 if $V - k_{x_i} + \varepsilon_{i0} > V - k_{x_0}$ where V is the patient valuation of treatment and ε_{i0} is a patient taste shock of traveling from x_i to x_0 which lies on the uniform distribution between 0 and $\bar{k}$. As $k_{x_0} = 0$ by construction, it follows that a patient located at x_0 would be willing to travel to x_i (that is, travel k_{x_i} minutes) with probability:

$$Pr(\varepsilon_{i0} > k_{x_i}) = \begin{cases} 1 - (1/\bar{k})k_{x_i} & \text{if} \quad k_{x_i} \leq \bar{k} \\ 0 & \text{if} \quad k_{x_i} > \bar{k} \end{cases} \tag{18}$$

We treat these probabilities as physician weights used to calculate the expected market share for a given location. This means the expected market share of a physician located at x_i for patients located x_0 is $E[S_{x_i}(x_0)] = \frac{Pr(\varepsilon_{i0} > k_{x_i})}{\sum_j Pr(\varepsilon_{j0} > k_{x_j})}$ where j indexes each physician in the database. For example, suppose there exist ten physicians all residing exactly at location x_0 while every other physician in the data resides over 80 minutes away. It follows that each of these ten physicians has equal probability of attracting patients from location x_0, resulting in each having an expected market share of 0.1 for patients located at x_0. It follows that the expected market share at location x_0 for physician *group* i that has N_i physicians in the group, located at x_i is $E[S^*_{x_i}(x_0)] = \frac{N_i Pr(\varepsilon_{i0} > k_{x_i})}{\sum_j N_j Pr(\varepsilon_{j0} > k_{x_j})}$.

As we have only county level information about where patients live in the SK&A data, we calculate an HHI for every geographic coordinate listed. Thus, we are in essence proxying patient location with the physician locations in the SK&A data. For each location, h, in the SK&A data, we calculate a distinct $HHI(x_h) = \sum_i E[S^*_{x_j}(x_h)]^2$ based on the expected market shares at location h. We then created an average concentration for the county as $HHI_c = \frac{1}{M_c} \sum_{h \in c} HHI(x_h)$ where M_c is the number of geographic points, h.

Finally, we merge the county-level HHI_c in the MarketScan® data. Since MarketScan® has information on the county of both the provider as well as the patient, we have information on where each physician's patients reside. For each physician, p, we take a weighted sum of the counties where physician p's patients reside to arrive at our physician level concentration measure $FTHHI^p_{phys} = \sum_c \omega_{cp} HHI_c$ where ω_c is the share of physician p's patients from county c.

B Variable Definitions

- Concentration Measures
  - $FTHHI_{phys}$: The Fixed-Travel-Time Herfindahl-Hirschman concentration measure for physicians. This measure is specific to each physician in the MarketScan® data. See Appendix A for details on construction.
  - HHI_{ins} : The Herfindahl-Hirschman concentration measure for insurance carriers. This measure is specific to each MSA. See Section 3 for details on construction.
- Expenditure Measures
  - P_n : The logarithm of price per service for episode of care n of services performed by the physician. See Section 3 for details on construction.
  - Q_n : The logarithm of service utilization for episode of care n provided by the physician. See Section 3 for details on construction.

- TE_n : The logarithm of total physician expenditures of episode of care n. See Section 3 for details on construction.

- Patient-Specific Controls (PAT)
  - $\ln(medinc_{pat})$ - The logarithm of the median income in the patient's county.
  - $educ_{pat}$ - The fraction of college educated individuals in the patient's county.
  - EPO - A dummy variable indicating if the patient's health plan is an exclusive provider organization.
  - HMO - A dummy variable indicating if the patient's health plan is a health maintenance organization.
  - POS - A dummy variable indicating if the patient's health plan is a point-of-service plan.
  - PPO - A dummy variable indicating if the patient's health plan is a preferred provider organization.
  - $HDHP$ - A dummy variable indicating if the patient's health plan is a high-deductible health plan.
  - $CDHP$ - A dummy variable indicating if the patient's health plan is a consumer-driven health plan.
  - $EMPLOYER$ - A dummy variable indicating if the patient's health plan is employer based
  - AGE - The patient's age
  - AGE^2 - The patient's age squared
  - AGE^3 - The patient's age cubed
  - $COMORBID$ - The number of co-morbidities (that is, concurrent diseases) of the patient.
  - $COMORBID^2$ - The number of co-morbidities squared.
  - $COMORBID^3$ - The number of co-morbidities cubed.
  - $GENDER$ - A dummy variable indicating the patient's gender

- Physician Quality Controls ($QUAL$)

- $\ln(medinc_{flow})$: The logarithm of the patient-weighted median household income. Here, $medinc_{flow} = \sum \omega_{cp} medinc_c$, where $medinc_c$ is the median income in county c and ω_{cp} is the share of physician p's patients from county c. Taken from the Area Resource File.
- $UNIV$: The fraction of hospitals in the physician's county, c, that are affiliated with a medical university. Taken from the Area Resource File.

- Physician Cost Controls ($COST$)
  - $\ln(rent_{phys})$: The logarithm of the median gross rent in the physician's county. Taken from the Area Resource File.
  - $\ln(medval_{phys})$: The logarithm of the median home value in the physician's county. Taken from the Area Resource File.
  - $\ln(medinc_{phys})$: The logarithm of the median household in the physician's county. Taken from the Area Resource File.
  - $\ln(facwage_{phys})$: The logarithm of the total health care facility payrolls divided by the number of facility employees. Taken from the Area Resource File.

- Population-Distribution Instruments (Pop. Dist)
  - $\ln(pop_{flow})$ - The logarithm of the patient-weighted total population. Here, $pop_{flow} = \sum \omega_{cp} pop_c$, where pop_c is the total population in county c and ω_{cp} is the share of physician p's patients from county c. Taken from the Area Resource File.
  - $\ln(pop65_{flow})$ - The logarithm of the patient-weighted population over 65 years of age. Here, $pop65_{flow} = \sum \omega_{cp} pop65_c$, where $pop65_c$ is the population over 65 in county c and ω_{cp} is the share of physician p's patients from county c. Taken from the Area Resource File.
  - $\ln(popK_{flow})$ - The logarithm of the patient-weighted population between K-10 and K years of age, where K = 65, 55, 45, and 35. Here, $popK_{flow} = \sum \omega_{cp} popK_c$, where ω_{cp} is the share of physician p's patients from county c and $popK_c$ is the imputed population between K and K - 10 years of age in county c in the Area Resource file. That is, $popK_c = fracK_c * pop_c$, where $fracK$ is the fraction in K to K-10 age-group for county c in the entire MarketScan® database.

- $\ln(pop_{MSA})$ - The logarithm of the population of the MSA. Taken from the Area Resource File.
- $URATE$ - The unemployment rate in the physician's county. Taken from the Area Resource File.
- $URATE^2$ - The unemployment rate in the physician's county squared. Taken from the Area Resource File.
- $URATE^3$ - The unemployment rate in the physician's county cubed. Taken from the Area Resource File.

- Firm-Distribution Instruments (Firm Dist.)
  - $\ln(firms_{flow})$ - The logarithm of the patient-weighted number of business establishments in the physician's county c in year y. Here, $firms_{flow} = \sum \omega_{cp} firm_c$, where $firm_c$ is the number of business establishments in the physician's county c and ω_{cp} is the share of physician p's patients from county c. Taken from U.S. Census Bureau's County Business Patterns database.
  - $\ln(firm20_{flow})$ - The logarithm of the fraction of business establishments with less than 20 employees in the physician's county c in year y. Here, $firm20_{flow} = \sum \omega_{cp} firm20_c$, where $firm20_c$ is the the fraction of business establishments with less than 20 employees in the physician's county c and ω_{cp} is the share of physician p's patients from county c. Taken from U.S. Census Bureau's County Business Patterns database.
  - $\ln(firm50_{flow})$ - The logarithm of the fraction of business establishments with greater than 20 employees and less than 50 employees in the physician's county c in year y. Here, $firm50_{flow} = \sum \omega_{cp} firm50_c$, where $firm50_c$ is the the fraction of business establishments with greater than 20 employees and less than 50 employees in the physician's county c and ω_{cp} is the share of physician p's patients from county c. Taken from U.S. Census Bureau's County Business Patterns database.
  - $\ln(firm100_{flow})$ - The logarithm of the fraction of business establishments with greater than 50 employees and less than 10 employees in the physician's county c in year y. Here, $firm100_{flow} = \sum \omega_{cp} firm100_c$, where $firm100_c$ is the the fraction of business establishments with greater than 50 employees and less than 100 employees in the physician's county c and ω_{cp} is the share of physician p's

patients from county c. Taken from U.S. Census Bureau's County Business Patterns database.

C Estimates of First-Stage Instrumental Variables

C.1 Population-Distribution Instruments

	Cardiology		Orthopedics	
	$\ln(FTHHI_{phys})$	$\ln(HHI_{ins})$	$\ln(FTHHI_{phys})$	$\ln(HHI_{ins})$
$\ln(pop_{flow})$	-0.047	0.330	1.401**	0.259
	(0.470)	(0.257)	(0.597)	(0.223)
$\ln(pop35_{flow})$	0.646**	-0.185	-0.909**	-0.204
	(0.293)	(0.166)	(0.380)	(0.144)
$\ln(pop45_{flow})$	-1.314***	-0.170*	-0.033	-0.068
	(0.284)	(0.095)	(0.255)	(0.087)
$\ln(pop55_{flow})$	0.812***	0.143**	-0.191	0.119**
	(0.169)	(0.065)	(0.167)	(0.056)
$\ln(pop65_{flow})$	-0.420***	-0.117**	-0.454***	-0.111***
	(0.141)	(0.048)	(0.143)	(0.043)
$\ln(pop_{msa})$	-0.196***	-0.050***	-0.195***	-0.052***
	(0.019)	(0.007)	(0.020)	(0.008)
$URATE$	-10.610	11.336*	18.540	17.714***
	(17.710)	(5.837)	(25.116)	(6.014)
$URATE^2$	-314.955	-510.975***	-526.274	-708.877***
	(518.731)	(168.950)	(699.779)	(178.567)
$URATE^3$	6365.101	4718.690***	5300.875	6184.862***
	(4268.093)	(1430.421)	(4875.836)	(1518.742)
F-Stat	69.2	13.1	60.4	22.3
Observations	3664391	3669352	4129905	4131962

Notes: The dependent variable is listed at the column head. Standard errors are clustered by provider-county. Not shown are the estimates on the covariates of specification (12). F-statistics test the null hypothesis that all instruments are jointly equal to zero. For each sample, the total effect of the unemployment rate on $\ln(HHI_{ins})$ (that is, $\widehat{\delta_1}URATE + \widehat{\delta_2}URATE^2 + \widehat{\delta_3}URATE^3$) is positive for all values within the 99th percentile unemployment rate.

C.2 Firm-Distribution Instruments

	Cardiology		Orthopedics	
	$\ln(FTHHI_{phys})$	$\ln(HHI_{ins})$	$\ln(FTHHI_{phys})$	$\ln(HHI_{ins})$
$\ln(firms)$	-0.440***	-0.024**	-0.307***	-0.033***
	(0.024)	(0.010)	(0.023)	(0.008)
$\ln(firm20)$	27.303***	7.721***	17.623***	7.379***
	(4.034)	(1.157)	(3.536)	(1.037)
$\ln(firm50)$	4.940***	1.310***	3.935***	1.293***
	(0.520)	(0.163)	(0.479)	(0.143)
$\ln(firm100)$	1.620***	0.327***	0.678***	0.281***
	(0.266)	(0.084)	(0.232)	(0.075)
F-Stat	168.5	28.3	95.9	39.7
Observations	3668971	673946	4133902	4135972

Notes: The dependent variable is listed at the column head. Standard errors are clustered by provider. Not shown are the estimates on the covariates of specification (12). F-statistics test the null hypothesis that all instruments are jointly equal to zero.

C.3 Instrument Validity

We report results of an exercise assessing the validity of the main instrument set. First we collect the residuals from specification (12) using the firm-distribution instrument set. Below, we run an OLS regression of these residuals on the population-distribution instrument set used in the study.

	Cardiology	Orthopedics
$\ln(pop_{flow})$	0.062	0.045
	(0.077)	(0.091)
$\ln(pop35_{flow})$	-0.042	0.009
	(0.049)	(0.060)
$\ln(pop45_{flow})$	0.012	-0.081
	(0.035)	(0.051)
$\ln(pop55_{flow})$	-0.030	0.043
	(0.024)	(0.031)
$\ln(pop65_{flow})$	-0.003	-0.012
	(0.021)	(0.021)
$\ln(pop_{msa})$	0.001	-0.004
	(0.003)	(0.004)
$URATE$	2.301	11.710***
	(2.659)	(3.664)
$URATE^2$	-26.729	-320.317***
	(78.132)	(108.009)
$URATE^3$	-261.935	1885.867**
	(602.158)	(782.913)
F-Stat	1.82	2.66
R^2	0.0003	0.0017
Observations	3664391	4129905

Notes: Standard errors are in parentheses and are clustered by provider. One, two, and three asterisks indicate significance at the 10-percent, 5-percent, or 1-percent significance level, respectively.

C.4 Additional Endogenous Controls

We provide additional estimates where we include controls for the size of the firm and the number of physicians per capita in the county. Specifically, we include a variable $\ln(scale)$ which is the logarithm of the average number of doctors per firm in county c at time t. We also include a variable $\ln(physdens)$ which is measured as the logarithm of the total number of cardiologists (or orthopedists) per capita in county c at time t. The former variable is meant to control for possible economies of scale of larger firms, while the latter variable is meant to control for the overall supply of physicians. As these two variables may be endogenous to the extent that physicians chase higher prices, we also include specifications where we include them as endogenous right-hand-side variables, which is labeled as "endogenous controls" in the tables.

All regressions include a dummy variable indicating the patient's gender, a polynomial of the patient's age (i.e. AGE, AGE^2, and AGE^3), a polynomial in the number of co-morbidities, as well as state-halfyear and disease/stage-of-illness fixed effects. The omitted plan types are "basic medical" and "comprehensive." Standard errors are in parentheses and are clustered by provider. One, two, and three asterisks indicate significance at the 10-percent, 5-percent, or 1-percent significance level, respectively.

C.4.1 Additional Endogenous Controls: Market Structure on Price

	Cardiology				Orthopedics			
	Exogenous Controls		Endogenous Controls		Exogenous Controls		Endogenous Controls	
$\ln(FTHHI_{phys})$	0.113***	0.110***	0.119***	0.062	0.097***	0.136***	0.107***	0.150**
	(0.016)	(0.018)	(0.016)	(0.048)	(0.019)	(0.035)	(0.023)	(0.076)
$\ln(HHI_{ins})$	-0.312***	-0.257**	-0.299***	-0.249	-0.200**	-0.406**	-0.295**	-0.428
	(0.094)	(0.126)	(0.103)	(0.237)	(0.087)	(0.174)	(0.132)	(0.639)
$\ln(scale)$	-0.060***	-0.059***	-0.070**	0.338	-0.033**	-0.061**	0.068	0.644**
	(0.013)	(0.012)	(0.033)	(0.216)	(0.014)	(0.024)	(0.082)	(0.310)
$\ln(physdens)$	0.044***	0.042***	0.013	-0.342	0.040***	0.062***	0.015	-0.317
	(0.009)	(0.008)	(0.030)	(0.240)	(0.011)	(0.019)	(0.036)	(0.380)
$\ln(medval_{phys})$	0.049**	0.052**	0.051**	0.156**	0.067***	0.059***	0.093***	0.310**
	(0.020)	(0.021)	(0.021)	(0.066)	(0.019)	(0.019)	(0.024)	(0.137)
$\ln(rent_{phys})$	-0.008	0.004	0.010	-0.237	-0.059	-0.052	-0.071	-0.041
	(0.058)	(0.057)	(0.067)	(0.214)	(0.044)	(0.050)	(0.058)	(0.201)
$\ln(facwage_{phys})$	-0.005	-0.002	0.013	0.040	0.005	-0.000	-0.011	-0.050
	(0.009)	(0.008)	(0.020)	(0.112)	(0.008)	(0.010)	(0.024)	(0.119)
$\ln(medinc_{phys})$	-0.061	-0.061	-0.063	-0.079	-0.097***	-0.133***	-0.140***	-0.384***
	(0.039)	(0.038)	(0.039)	(0.068)	(0.035)	(0.044)	(0.047)	(0.133)
$\ln(medinc_{flow})$	0.007	0.013	0.001	-0.172	0.099***	0.113***	0.100***	0.090
	(0.022)	(0.021)	(0.023)	(0.109)	(0.018)	(0.025)	(0.021)	(0.088)
UNIV	0.018	0.020	0.029*	0.004	-0.000	-0.012	-0.020	-0.073
	(0.014)	(0.013)	(0.017)	(0.083)	(0.012)	(0.018)	(0.025)	(0.132)
$\ln(medinc_{pat})$	0.041***	0.042***	0.042***	0.017	0.027***	0.030***	0.027***	0.027
	(0.008)	(0.008)	(0.009)	(0.022)	(0.008)	(0.010)	(0.009)	(0.022)
$\ln(educ_{pat})$	0.112***	0.098***	0.138***	0.350	0.151***	0.171***	0.157***	0.290
	(0.032)	(0.037)	(0.037)	(0.217)	(0.030)	(0.038)	(0.039)	(0.293)
EPO	-0.032***	-0.029***	-0.031***	-0.027**	-0.025***	-0.014	-0.019**	0.003
	(0.007)	(0.007)	(0.007)	(0.011)	(0.008)	(0.011)	(0.009)	(0.030)
HMO	-0.038***	-0.037***	-0.038***	-0.050***	-0.005	-0.005	-0.005	-0.004
	(0.006)	(0.006)	(0.006)	(0.012)	(0.006)	(0.007)	(0.007)	(0.013)
POS	-0.018***	-0.017***	-0.018***	-0.020*	-0.004	-0.006	-0.007	-0.020*
	(0.005)	(0.005)	(0.005)	(0.011)	(0.005)	(0.006)	(0.006)	(0.011)
PPO	-0.000	0.001	-0.000	0.004	0.007*	0.005	0.006	0.002
	(0.004)	(0.004)	(0.004)	(0.007)	(0.004)	(0.004)	(0.004)	(0.008)
HDHP	-0.003	-0.003	-0.002	0.002	0.004	0.005	0.005	0.009
	(0.012)	(0.012)	(0.013)	(0.013)	(0.007)	(0.007)	(0.007)	(0.012)
CDHP	0.030***	0.030***	0.030***	0.026***	0.038***	0.038***	0.037***	0.034***
	(0.005)	(0.005)	(0.005)	(0.008)	(0.004)	(0.005)	(0.005)	(0.009)
EMPLOYER	-0.005	-0.005	-0.004	-0.008	0.024***	0.023***	0.022***	0.010
	(0.009)	(0.009)	(0.009)	(0.011)	(0.006)	(0.007)	(0.007)	(0.013)
Instruments	Pop. Dist.	Firm Dist.	Pop. Dist.	Firm Dist.	Pop. Dist.	Firm Dist.	Pop. Dist.	Firm Dist.
Observations	3648616	3653195	3648616	3653195	4115084	4119081	4115084	4119081

C.4.2 Additional Endogenous Controls: Market Structure on Utilization

	Cardiology				Orthopedics			
	Exogenous Controls		Endogenous Controls		Exogenous Controls		Endogenous Controls	
$\ln(FTHHI_{phys})$	0.106**	0.151***	0.085*	0.059	0.005	0.021	0.027	0.039
	(0.044)	(0.048)	(0.050)	(0.080)	(0.025)	(0.027)	(0.039)	(0.043)
$\ln(HHI_{ins})$	-0.398*	-0.211	-0.397	-0.379	-0.114	-0.159	-0.205	-0.017
	(0.221)	(0.285)	(0.303)	(0.399)	(0.107)	(0.126)	(0.203)	(0.376)
$\ln(scale)$	-0.074***	-0.237***	-0.059**	-0.184	0.016	0.077	0.001	0.129
	(0.028)	(0.081)	(0.030)	(0.370)	(0.019)	(0.090)	(0.028)	(0.223)
$\ln(physdens)$	0.025	-0.149**	0.022	0.523	0.011	-0.047	0.021	-0.258
	(0.024)	(0.072)	(0.025)	(0.336)	(0.014)	(0.036)	(0.021)	(0.245)
$\ln(medval_{phys})$	-0.058*	-0.065*	-0.062*	-0.127	-0.065***	-0.033	-0.067***	0.049
	(0.030)	(0.039)	(0.033)	(0.080)	(0.018)	(0.025)	(0.019)	(0.097)
$\ln(rent_{phys})$	0.450***	0.654***	0.420***	0.425	0.101*	0.117*	0.110*	0.205
	(0.114)	(0.162)	(0.129)	(0.367)	(0.055)	(0.065)	(0.057)	(0.135)
$\ln(facwage_{phys})$	-0.024	0.118**	-0.028	-0.218	-0.011	-0.010	-0.013	0.030
	(0.023)	(0.055)	(0.021)	(0.156)	(0.010)	(0.024)	(0.011)	(0.072)
$\ln(medinc_{phys})$	-0.031	-0.036	-0.015	0.008	0.084**	0.042	0.063	-0.013
	(0.076)	(0.087)	(0.071)	(0.094)	(0.042)	(0.052)	(0.048)	(0.102)
$\ln(medinc_{flow})$	-0.100*	-0.110**	-0.126**	0.063	-0.037	-0.039	-0.026	-0.067
	(0.053)	(0.053)	(0.053)	(0.167)	(0.026)	(0.028)	(0.029)	(0.059)
UNIV	0.018	0.115**	0.014	-0.078	-0.033**	-0.036	-0.039*	0.003
	(0.036)	(0.049)	(0.037)	(0.121)	(0.016)	(0.029)	(0.021)	(0.084)
$\ln(medinc_{pat})$	0.080***	0.092***	0.076***	0.087***	0.024**	0.025**	0.027**	0.024
	(0.019)	(0.023)	(0.019)	(0.033)	(0.011)	(0.012)	(0.012)	(0.015)
$\ln(educ_{pat})$	-0.203**	-0.052	-0.196**	-0.629*	-0.329***	-0.290***	-0.321***	-0.141
	(0.087)	(0.115)	(0.099)	(0.323)	(0.043)	(0.050)	(0.048)	(0.189)
EPO	0.037**	0.038**	0.029	0.029	0.009	0.013	0.013	0.010
	(0.018)	(0.018)	(0.018)	(0.021)	(0.011)	(0.012)	(0.013)	(0.020)
HMO	0.009	0.014	0.009	0.017	-0.025***	-0.024***	-0.025***	-0.020*
	(0.016)	(0.016)	(0.016)	(0.022)	(0.009)	(0.009)	(0.009)	(0.011)
POS	0.018	0.013	0.018	0.029	0.011	0.008	0.010	0.006
	(0.018)	(0.019)	(0.019)	(0.022)	(0.007)	(0.008)	(0.008)	(0.009)
PPO	0.016	0.014	0.015	0.018	-0.007	-0.008	-0.008	-0.007
	(0.013)	(0.013)	(0.013)	(0.015)	(0.006)	(0.006)	(0.006)	(0.008)
HDHP	0.047**	0.057**	0.046**	0.034	0.012	0.013	0.013	0.014
	(0.021)	(0.024)	(0.021)	(0.025)	(0.014)	(0.015)	(0.015)	(0.015)
CDHP	0.016	0.021	0.015	0.014	-0.038***	-0.038***	-0.038***	-0.036***
	(0.016)	(0.017)	(0.016)	(0.018)	(0.008)	(0.008)	(0.008)	(0.009)
EMPLOYER	0.041***	0.048***	0.041***	0.035*	-0.042***	-0.044***	-0.042***	-0.046***
	(0.015)	(0.015)	(0.015)	(0.020)	(0.007)	(0.007)	(0.007)	(0.009)
Instruments	Pop. Dist.	Firm Dist.	Pop. Dist.	Firm Dist.	Pop. Dist.	Firm Dist.	Pop. Dist.	Firm Dist.
Observations	3648616	3648616	3653195	3653195	4115084	4115084	4119081	4119081

D Switching Regression with Out-of-Pocket Price in Demand Equation

As an additional specification, we attempt to improve the switching regression by conforming more closely to the theoretical model. According to our theory, patients and physicians respond to distinct prices. Patients respond to the out-of-pocket price while

Table 7: Switching Regression with Out-of-Pocket Price in Demand Equation

	Cardiology		Orthopedics	
γ^D	-0.773***	-1.01***	-0.482***	-0.710***
	(0.064)	(0.098)	(0.051)	(0.077)
γ^S	1.066***	0.449*	0.403**	0.314*
	(0.334)	(0.268)	(0.181)	(0.155)
$Pr(\omega_n = 1)$	0.276	0.212	0.478	0.397
Instruments	Pop. Dist.	Firm Dist.	Pop. Dist.	Firm Dist.
Log Likelihood	-4.64e6	-4.61e6	-6.71e6	-7.15e6
Observations	2962323	2962323	3796120	3796120

Notes: Standard errors are in parentheses and are clustered by provider. One, two, and three asterisks indicate significance at the 10-percent, 5-percent, or 1-percent significance level, respectively.

physicians respond to the service price. To account for this dichotomy, we adjust the patient demand equation by replacing the service price, P_n, with the out-of-pocket price, $P^{pock} = \alpha_n P_n$. This alteration in the specification enhances identifying power by creating two distinct functional forms for each regime. It also uses a more accurate measure of the out-of-pocket price. Of course, this specification change also comes at the expense of including additional endogenous variation. Not only do the instruments have to be valid in terms of determining the negotiated service price, P_n^*, but they must also be valid in terms of determining the benefits chosen, α_n^*. The instruments will be strong if insurance carrier entry and exit, occurring in Period 0, helps determine the level of benefits, which was demonstrated by the estimates in Table 3. The instruments will be invalid if unobservable variables that affect utilization of the individual are systematically correlated with the population (or firm size) distribution and benefit selection.[72]

The results in Table 7 and Table 5 are qualitatively very similar, showing upward sloping physician supply and downward sloping patient demand. However, the magnitude of the elasticities implied by the estimates are quite distinct. Of note is that the estimates of the demand elasticities are considerably larger (in absolute value). This result may be attributable to better identification as well as the result of using more accurate variation

[72]For example, a bias may arise if there are unobserved health conditions that are related to both the population distribution and insurance selection. This type of bias is unlikely given the detailed individual level information included in the analysis. Moreover, this bias is especially unlikely using the firm size distribution instruments, which is arguably less related to the health of the population.

in the out-of-pocket price. The fact that we observe some differences in the magnitudes across these two empirical models should not be surprising, since the two empirical models are quite distinct. Here we are estimating an elasticity with respect to the out-of-pocket price, while in Table 5 we are reporting elasticities with respect to a service price.

To compare the patient demand elasticity estimates to other estimates in the literature, we need to calculate the demand elasticity for the full population of patients. The out-of-pocket price elasticity for the full population of patients may be calculated by multiplying the patient's elasticity by the fraction of observations where the patient's constraint is binding. For example, the elasticities for cardiology are -0.213 (=-0.773·0.276) and -0.214 (=-1.01·0.212) and the elasticities for orthopedics are -0.230 (=-0.482·0.478) and -0.282 (=-0.710·0.397). These estimates are consistent with the out-of-pocket elasticity estimates measured using randomized data from the RAND health insurance experiment that finds elasticities in the -0.1 to -0.2 range.[73]

[73] We also ran a linear IV specification where we included both $\ln(P)$ and $\ln(P^{pock})$ in the same specification: $\ln(Q_n) = \gamma_1 \widehat{\ln(P_n)} + \gamma_2 \widehat{\ln(P_n^{pock})} + \delta' COST + \kappa' QUAL + \lambda' PAT + \zeta_{at} + \zeta_d + \varepsilon_n$. The coefficient γ_1 provides an estimate of the marginal effect of a change in service price on service utilization holding fixed the out-of-pocket price and can therefore be interpreted as an estimate of the physician's supply elasticity. The coefficient on γ_2 provides the marginal effect of a change in out-of-pocket price, holding fixed any supply response due to variation in the service price and can be interpreted as an estimate of the demand elasticity. Using the population distribution insturments, the supply estimates were 1.00 (0.156) for cardiology and 0.286 (0.067) for orthopedics. Demand elasticity estimates with respect to out-of-pocket price were -0.402 (0.084) and -0.199 (0.047), respectively. However, identifying both the physician and patient responses to price in this linear framework proved to be more difficult using the firm-size distribution instruments. Applying these alternative instruments, we found both the demand and supply elasticities to be insigificant.