Insurer-Provider Networks in the Medical Care Market*

Katherine Ho

Harvard University

kho@fas.harvard.edu

Preliminary Version - Please do not quote

October 1, 2004

Abstract

Managed care health insurers in the US restrict their enrollees' choice of hospitals to specific networks. This paper investigates the causes and welfare effects of the observed hospital networks. A simple profit maximization model explains roughly half the observed contracts between insurers and hospitals. A generalization of the model demonstrates an additional effect: hospitals that do not need to contract with all insurance plans to secure demand (for example, hospitals that are capacity constrained under a limited or selective network) can force insurers to compete for contracts. Some plans may exclude these hospitals in equilibrium. Hospitals can merge to form "systems", which may also affect bargaining between hospitals and insurance plans. I estimate the expected division of profits between insurance plans and different types of hospitals using data on insurers' choices of network. Hospitals in systems are found to capture around $179,000 of incremental profits per month compared to other providers and to impose high penalties on plans that exclude their partners. Providers that are expected to be capacity-constrained capture an additional $1800 per patient on average. I show that these high markups imply a negative incentive for hospitals to invest to remove capacity constraints, despite a median benefit to consumers of over $330,000 per new bed per year.

*I am grateful to Ariel Pakes and David Cutler for their advice and encouragement. I also thank John Asker, Nancy Dean Beaulieu, Oliver Hart, Joy Ishii, Julie Mortimer, Michael Ostrovsky and seminar participants at Harvard University for helpful comments and suggestions. All remaining errors are my own.

1 Introduction

The effects of managed care health insurers on the price and quality of medical care have been widely researched[1]. One aspect of their impact, however, has not been addressed in detail: the restriction imposed by each insurance plan on the network of hospitals from which its enrollees can choose. In a previous paper (Ho, 2004) I estimate that a move from the existing hospital networks to an unselective outcome where all plans offer a free choice of hospitals would lead to a total gain in consumer surplus of $1.04 billion per year and a loss to producers of just $0.80 million per year. In this paper I demonstrate a second important consequence of the process used to determine insurers' contracts with hospitals: an under-investment in capacity by hospitals that expect to fill their beds under a limited or selective network. I show that investment in new beds would generate a benefit to consumers of over $330,000 per bed per year assuming fixed premiums, far outweighing the effect on hospital and insurer profits. However, the bargaining process by which the prices paid to providers are agreed implies a negative incentive for capacity-constrained hospitals to invest.

I base my analysis on a dataset that defines the network of every managed care plan in 43 markets across the US. On average 10 per cent of potential insurer-hospital pairs in my data do not arrange contracts to provide care. The proportion varies from zero per cent in some markets to as many as 50 per cent in others. I define selective markets as those in which at least four of the five major plans fail to reach agreement with at least one major hospital: by this definition roughly 20 per cent of observed markets are selective.

I use demand estimates from my previous paper to calculate the producer surplus (defined as plan revenues less hospital costs of care) generated by each potential hospital network for each plan in the data, taking into account patient flows across plans and hospitals. A simple analysis shows that around 50 per cent of contractual decisions are explained by this definition of producer surplus. A model of the price-setting negotiation between insurers and providers is needed to explain the rest of the data. I set out a simple motivating example to demonstrate how the negotiation can influence contractual decisions, even holding surplus fixed. I show that hospitals that do not need to contract with all plans in order to secure demand, such as those that expect to be capacity constrained even under a limited or selective network, can require plans to compete for contracts. This competition drives up prices; as a result, some

[1] For example, Miller and Luft (1997) review fifteen studies of the effects of managed care on quality. They find no compelling evidence of a reduction in quality of care, although patients show less satisfaction with managed care than with traditional plans. Cutler, McClellan and Newhouse (2000) consider the causes of the expenditure reductions achieved by managed care plans in the treatment of heart disease. They show that virtually all the difference in spending between indemnity plans and HMOs comes from lower unit prices rather than the quantity of services or a difference in health outcomes.

plans exclude these hospitals in equilibrium. In addition, if hospitals merge to form systems they can both demand higher prices than other providers and effectively force plans to contract with all or none of the hospitals in the system. Reduced form analysis is consistent with the model's predictions.

The model also predicts welfare effects. I consider the example of capacity-constrained hospitals in detail, and show that the predicted selective equilibrium can be inefficient even when the excluded hospital is full at equilibrium and the consumers with the highest value for it are the ones treated. The inefficiency is generated because consumers are forced to make sub-optimal choices across insurers in order to access their preferred hospital. The resulting loss of consumer welfare may outweigh the benefit derived when the highest-valuation patients are given preferential access to the hospital.

I estimate the profits secured by different types of hospital using insurers' observed choices of hospital networks and data on the characteristics of both insurers and providers. The analysis is complicated by the fact that each health plan's choice of network affects the optimal decisions of all other plans in its market. This mutual dependence of firms' decisions leads to the existence of multiple potential equilibria and to possible problems with endogenous regressors. A number of existing papers have estimated models that address these issues[2]; however, their approaches are feasible only for problems involving small numbers of firms. This paper adopts a different approach presented in Pakes, Porter, Ho and Ishii (2004) in which plans choose their networks in a simultaneous-moves game, conditional on their expectations regarding other plan choices and the prices that they predict will be determined through bargaining. The equilibrium implicitly establishes markups for a hospital's services that are functions of the characteristics of the hospital itself and the distribution of consumer, hospital, and plan characteristics in the particular market[3].

I estimate the markups of three specific hospital types: those that expect to be capacity-constrained, "star" hospitals (providers that are very attractive to consumers) and those that are members of hospital systems. I find that, as predicted by the theory model, hospitals that can credibly threaten to turn down low price offers capture a high share of the surplus. Hospitals with exogenous characteristics predicting capacity constraints capture $1800 per patient treated more than other providers. The markups of hospitals in systems are approximately $179,000 per month higher than other providers; they also charge high penalties from plans that contract with some but not all of the hospitals in their system. The results are therefore consistent with several recent papers that suggest that hospital mergers may prevent plans

[2] Seim (2001), Andrews, Berry and Jia (2004) and Ciliberto and Tamer (2003) are examples.

[3] The identifying assumption used for estimation is the simple necessary condition that the profit each plan expects to earn from its chosen network is greater than the profit it could expect to earn from alternative networks. This assumption is sufficient to define bounds on the feasible values of the parameters to be estimated.

from using the threat of exclusions to control prices[4]. Star hospitals may also capture high profits but these are imprecisely estimated. In addition, I find that hospitals with low costs have higher markups than their competitors, consistent with simple bargaining models.

Finally, I outline the implications of the results. While the overall picture implied by the data consistently points to the effects outlined above, the estimated confidence intervals indicate significant variance about the point estimates, making statements about precise magnitudes difficult. In addition, the methodology used in this paper does not fully detail how the multiplicity of possible equilibria is resolved. Instead my analysis provides a reduced form characterization of the markup equation that is generated by the equilibrium that does materialize. More precise details of the bargaining game would be needed to determine how that equilibrium is chosen, and this would be required before I could provide a detailed analysis of counterfactuals. I do, however, analyze the relationship between hospital characteristics and markups. This provides helpful information for assessing which bargaining models best describe the hospital-insurer price negotiation. To the extent that policy and environmental changes do not affect the reduced form relationship, the results can also be used to predict how changes in market characteristics are likely to affect hospital markups.

The welfare effects implied by the estimates are considered in this spirit. In particular I predict the impact of hospital investment to remove capacity constraints, assuming no effect of this change on plan networks or other market characteristics. The investment would result in a median benefit to consumers of $0.20 per person per new bed per year, or $338,800 per year per additional bed for the market as a whole. Insurer profits would increase as a result of the change. However, I find that the hospital markups generated by capacity constraints outweigh the additional revenues from new patients and imply that, at least in the short term, these providers have no incentive to invest.

Several strands of the health economics literature are relevant to this paper. A number of authors demonstrate that the prices paid by plans to hospitals are consistent with simple bargaining models[5]. Gal-Or (1997, 1998) develops a Nash bargaining model in a two-plan, two-hospital setting. Vistnes (2000) has a model of two-stage competition between hospitals: providers compete first for preferential acesss to health plans and then for individual patients. Finally, Eggleston, Norman and Pepall (2002) use a similar theoretical framework in a market containing health plans, hospitals and physician groups to look at the effects of horizontal and vertical integration on prices. However, previous authors have

[4] See, for example, Lesser and Grinsburg (2001), Mays, Hurley and Grossman (2003), and Capps and Dranove (2004).

[5] Most of these regress the prices paid to hospitals on measures of hospital and plan bargaining power. Examples are Brooks, Dor and Wong (1996), Zwanziger and Mooney (2000) and Feldman and Wholey (2001). In addition, Town & Vistnes (2001) and Capps, Dranove and Satterthwaite (2004) both investigate the effect of the hospital's value to consumers on its profits. They estimate consumer preferences over hospitals and regress hospital profits or prices on variables that summarize consumer demand for the hospital.

not addressed the network allocation question empirically: that is the main contribution of this paper.

In the next two sections I describe the contractual process between insurers and hospitals and introduce the dataset. Section 4 outlines the demand estimates from Ho (2004) and uses them to generate a measure of surplus. Section 5 sets out the simple theory model of bargaining; Section 6 contains reduced form results; and Section 7 introduces the full empirical model. Results and counterfactuals are given in Sections 8 and 9, and the final section concludes.

2 Industry Background

Each year, every privately insured consumer in the US chooses a health plan, generally from a menu offered by his employer[6]. The insurer contracts with hospitals and physicians to provide any care needed during the year. When the consumer requires medical care, he may visit any of the providers listed by the health plan, and receive services at zero charge or after making a small out-of-pocket payment.

There is some variety in the restrictiveness of different types of managed care plan. If an individual is insured by a Health Maintenance Organization (HMO) he may visit only the hospitals in that plan's network. Point of Service (POS) plan enrollees can visit out-of-network hospitals but only if referred by a Primary Care Physician. Preferred Provider Organizations (PPOs) and indemnity plans are the least restrictive insurers: enrollees do not need a PCP referral to visit an out-of-network hospital, although PPOs may impose financial penalties for doing so, for example in the form of increased copayments or deductibles. The focus of this paper is on HMO and POS plans, since their network choices have the strongest effect on both consumers and hospitals; 53 per cent of the privately-insured population was enrolled in an HMO/POS plan in 2000.

2.1 The Contractual Process

Every HMO/POS plan contracts separately with every hospital in its network. The exact form of the contracts varies, but all specify a price to be paid to the provider per unit of care (for example a price per inpatient day or per diagnosis-related group (DRG)). Prices vary both across providers for a given insurer and across insurers for a given provider; contracts are usually renegotiated annually. Both parties in the negotiation need to balance consumer demand for services against the price agreed. A health plan would prefer to contract with the hospitals that are valued by its likely customers, particularly the customers on the margin of joining, but must also take into account the fact that hospitals in demand may seek

[6] 57 per cent of the population is insured through an employer, compared to 5 per cent who purchase insurance independently and 24 per cent in Medicare and Medicaid. (See the website www. statehealthfacts.org.)

higher prices than their less differentiated counterparts. Hospitals seek to maximize their returns by contracting with plans that both offer high prices and provide a steady flow of patients.

Interviews with plan and hospital representatives who are involved in contractual negotiations confirm that the form of the price negotiation, and therefore the effect of contracting decisions on final prices, is affected by the relative bargaining positions of hospitals and health plans in the market. As the Director of Operations Analysis in one hospital chain put it:

> "There are counteracting effects here: the outcome [of any plan decision, like excluding a particular hospital] depends on where the balance of power lies."

In many markets, where managed care is strong and hospitals compete for contracts, the plan may begin negotiations by stating the maximum per cent increase in prices it can offer from the previous year and hospitals may generally accept that price increase. The Executive Director of another hospital system described one potential outcome in such markets:

> "There are examples where there were too many hospitals in an area, and the plans played them off against each other to the point where the price paid was no more than marginal cost."

This idea is consistent with the theory that, when plans have better outside options than hospitals, they are able to exclude those that demand high prices. The negotiation can be very different in markets where a few hospitals have very strong reputations and high market shares: the hospital may be able to demand a price which the plan must pay in order to avoid losing the contract[7]. A hospital Director said the following:

> "In market X [where hospitals are very strong], the prices [the best hospitals] charge are based on their very high patient satisfaction results and their strong reputation. They can get high prices from any plan in the market, and they don't need them all."

The CEO of a small hospital in a different market had a similar story:

> "Large [hospitals] in this market can dictate whatever prices they want. The bigger names can demand the highest prices."

[7] Prices paid to hospitals were regulated at the state level in the 1960s and 1970s. However, since Medicare and Medicaid switched from cost-based to prospective payment systems, and managed care encouraged increased price competition between hospitals, rate regulation has virtually disappeared.

The implication is that, in markets where hospitals expect to be full or can credibly threaten to turn down low price offers for some other reason, they demand a large share of the surplus they generate; this may prompt plans to exclude them.

2.2 The Timing of Firm Decisions

In order to model the contractual process I need to specify the timing of the different hospital and plan decisions and make a number of other assumptions. These are summarized in the following section.

The stages of my model are as follows:

Stage 1: Plans choose their negotiating partners; plan-hospital pairs agree on prices

Stage 2: Plans set premiums

Stage 3: Consumers and employers jointly choose plans

Stage 4: Sick consumers visit hospitals; plans pay hospitals per service provided

My main focus is on Stage 1. I assume fixed premiums throughout most of my calculations; I include a robustness test to consider the effects of potential premium adjustments in Section 8. I analyze Stages 3 and 4 in Ho (2004): my methodology is outlined in section 4.1 and the results of that study are incorporated where necessary in this paper.

A few additional comments are in order. First, many insurers offer several types of product: for example Aetna health plan may offer both an HMO and a POS product in a given market. I assume that the choice of products, together with the hospital's choices of location, services, quality and capacity, are made prior to Stage 1. My analysis conditions on these decisions. I therefore do not explicitly model issues such as product-based price discrimination (the plan's choice of products can be seen as a way of dividing the market into segments with different price elasticities of demand) and the hospital's decision regarding investment in specific services given those offered by its competitors. Similarly, I assume that hospital merger decisions are made prior to the contractual process[8]. Second, I model plan choices as being independent across markets: that is, I assume that Aetna in Boston chooses its network to maximize its profits in that market alone; Aetna in San Francisco performs a similar market-specific profit maximizing calculation. Third, the model gives health plans the power to choose the hospitals with which they negotiate; hospitals are excluded from this decision. However, the theories discussed

[8] These assumptions seem reasonable because the relevant variables change more slowly over time than hospital-insurer contracts. For example, over 90% of hospitals did not alter their offerings of angioplasty, ultrasound, open heart surgery or neonatal intensive care units over the four-year period 1997-2001; 70 percent of hospitals changed their capacity levels by fewer than 20 beds over the same four-year period. Plan product offerings and hospital locations are similarly static. Hospital-insurer contracts, in contrast, are usually renegotiated annually. My goal is to estimate the static equilibrium in contracts within the dynamic game in which these longer-run choices are made.

later in the text make clear that this does not prevent hospitals from capturing up to 100 per cent of the surplus in some cases: hospitals are free to refuse plan price offers so the equilibrium outcome depends on their preferences as well as those of insurers.

My dataset contains no exclusive contracts (either hospitals reaching agreement exclusively with a single insurer or vice versa), and few vertically integrated organizations. Many hospitals and insurers attempted vertical integration in the 1990s but this has become increasingly rare in recent years; the literature implies that the breadth of skills needed to run both a hospital and a plan is too large for the vertically integrated model to be viable except in very specific circumstances[9]. The key exception to this pattern is Kaiser Permanente, a dominant HMO in California and elsewhere that owns a large number of hospitals. Since vertical integration seems to be disappearing I do not attempt to explain the phenomenon in this paper. I condition on the existence of Kaiser health plans and hospitals in my analysis of both the supply and demand sides of the market (since they are important members of the plan and hospital choice sets, particularly in California) but exclude them from my models of firm behavior.

The health plan must take state and federal legislation into account when choosing its providers. Many states have implemented Any Willing Provider laws which prohibit health insurers from excluding qualified health care providers that are willing to accept the plans' terms and conditions. However, these regulations have been argued to remove the benefits of managed care, since they prevent plans from trading volume for lower provider prices. Perhaps for this reason they apply to hospitals in only seven states (in other areas they are largely limited to pharmacies). I have data covering two markets within these states; I find that plans are just as likely to exclude hospitals in these markets as elsewhere. I therefore assume that these regulations have no impact on plan decisions in the markets I consider. In addition, some states have implemented Essential Community Provider laws, which require insurers to contract with providers that offer "essential community services", such as public hospitals and teaching hospitals, and to contract with enough hospitals to serve the needs of the local population. I assume these regulations do not affect the decision of a plan to exclude any particular hospital since consumer demand forecasts would prevent it from dropping too many hospitals in any case.

3 A Preview of the Dataset

The primary dataset analyzed in this paper defines, for every HMO/POS plan in 43 major US markets, the network of hospitals offered to enrollees in March/April 2003. The dataset includes 516 HMO/POS

[9] See, for example, Burns and Pauly (2002) and Burns and Thorpe (2000)

plans and 665 hospitals in total. The information was collected from individual plan websites; missing data were filled in by phone[10]. Figure 1 documents the significant variation across both markets and plans in the extent to which plans exclude major hospitals from their networks. Markets are categorized on a scale from 1 to 5, where 1 is the least selective, indicating that each of the five largest plans (by enrollment) contracts with all eight largest hospitals (by number of admissions). In markets ranked 5, at least four of the largest plans exclude at least one major hospital; the other categories lie between these extremes. Markets are fairly evenly spread across the five categories: 16 markets are ranked 1 or 2 (not selective) and 21 are ranked 4 or 5 (very selective). The figure also shows the distribution of plans by the number of major hospitals excluded, and the variation in this distribution across types of market. Plans' selective behavior varies widely: 217 plans exclude no major hospitals, but 62 plans exclude at least four of the eight major hospitals in their markets.

Table 1 compares the means of a number of market characteristics in selective and unselective markets. There are few significant differences. Selective markets do not have significantly smaller populations, more hospitals, or more beds per capita than unselective markets and are not clustered geographically. There are no significant demographic differences. The only difference that is significant at p=0.05 (or in fact at p=0.2) is the standard deviation of the distances between hospitals in the market. Plans seem to be more willing to exclude hospitals in areas where hospitals are clustered into several groups, perhaps because each provider in a given group is a reasonable substitute for the others. The raw data therefore do not offer an obvious explanation for the observed variation; however, they do provide a hint that demand effects may be important. These are taken into account in the analysis described in Section 4.

A number of other datasets are introduced at various stages of the analysis. Hospital characteristic data is taken from the American Hospital Association (AHA) dataset for 2001. Plan characteristics come from two datasets from Atlantic Information Services[11] for Quarters 3 and 4 of 2002, supplemented with information from the *Weiss Ratings' Guide to HMOs and Health Insurers* for Fall 2002. Data on plan performance comes from the *Health Employer Data and Information Set* (HEDIS) and the *Consumer Assessment of Health Plans* (CAHPS) 2000, both of which are published by the National Committee for Quality Assurance (NCQA). These data measure clinical performance and patient satisfaction in 1999. My previous paper, Ho (2004), uses all these datasets: further details on the data, and the methodology

[10] The markets are: Atlanta GA, Austin TX, Baltimore MD, Boston MA, Buffalo NY, Charlotte NC, Chicago IL, Cincinnati OH, Cleveland OH, Columbus OH, Dallas TX, Denver CO, Detroit MI, Fort Worth TX, Houston TX, Indianapolis IN, Jacksonville FL, Kansas City MO, Las Vegas NV, Los Angeles CA, Miami FL, Milwaukee WI, Minneapolis MN, New Orleans LA, Norfolk VA, Oakland CA, Orange County CA, Orlando FL, Philadelphia PA, Phoenix AZ, Pittsburgh PA, Portland OR, Sacramento CA, St. Louis MO, Salt Lake City UT, San Antonio TX, San Diego CA, San Francisco CA, San Jose CA, Seattle WA, Tampa FL, Washington DC, and West Palm Beach FL.

[11] These are *The HMO Enrollment Report* and *HMO Directory 2002*. Both are based on plan state insurance filings.

used to create additional variables (such as plan market shares) that are employed again in this analysis, are given there. My demand estimation includes all 665 hospitals and all 516 managed care plans in the data. When I consider the supply side I restrict attention to non-Kaiser plans for which premiums are observed; I also exclude a few extremely selective insurers that I regard as outliers[12]. The remaining data contains 451 plans in total[13]. I model these plans' contracts with the six largest hospitals in each market: these cover an average of 57 per cent of the total admissions to non-Federal general medical and surgical hospitals in the markets I consider. I condition on the observed contracts between each plan and the remaining, smaller hospitals in the market.

4 Effect of the Network on Total Plan and Hospital Profits

4.1 Demand Estimates

In order to understand the equilibrium network outcomes I need to analyze Stages 3 and 4 of the model, in which consumers choose their health plans taking into account the hospitals they expect to visit in the coming year. The parameter estimates generated in Ho (2004) are used as an input to this paper's supply side analysis. The demand estimation process has three stages:

1. The first step is to estimate demand for hospitals using a discrete choice model that allows for observed differences across individuals. With some probability consumer i (whose type is defined by age, sex and zipcode tabulation area (ZCTA)) becomes ill. His utility from visiting hospital h given diagnosis l is given by:

$$u_{ihl} = \eta_h + x_h\alpha + x_h\upsilon_{il}\beta + \varepsilon_{ihl}$$

   where x_h, η_h are vectors of observed and unobserved hospital characteristics respectively, υ_{il} are observed characteristics of the consumer such as diagnosis and location and ε_{ihl} is an idiosyncratic error term assumed to be iid Type 1 extreme value. Hospital characteristics include location; the number of beds; the numbers of nurses and doctors per bed; and details of services offered, owner-

[12] I exclude plans that drop more than four of the top six hospitals because these may have different reasons for their contracting decisions than other plans in the data. I also exclude two specific outliers: Scott and White Health Plan of Austin, TX and Group Health Cooperative of Puget Sound. These are different from most other plans in the market in that they are locally-based, consumer-driven insurers that are heavily focused on primary care.

[13] The supply side analysis includes a prediction of plan market shares given the networks offered by every plan in the market. I condition on the existence of "excluded" plans (Kaiser plans, those that are very selective and those for which premiums are unobserved) when calculating the shares of the plans that are modelled explicitly. I also take account of indemnity and PPO plans, making assumptions about their characteristics, and allow consumers to choose the outside option of being uninsured, as described in detail in Ho (2004).

ship and accreditation. This equation is estimated using standard maximum likelihood techniques and micro (encounter-level) data from the MEDSTAT MarketScan Research Database for 1997-98. The data provide information on the hospital admissions of indemnity plan and PPO enrollees[14].

2. Secondly, I use the estimated coefficients to predict the utility provided by each plan's hospital network. Individual i's expected utility from the hospital network offered by plan j in market m is calculated as:

$$EU_{ijm} = \sum_{l} p_{il} \log \left(\sum_{h \epsilon H_j} \exp(\eta_h + x_h \hat{\alpha} + x_h \nu_{il} \hat{\beta}) \right)$$

where p_{il} is the probability that individual i will be hospitalized with diagnosis l.

3. Finally, I use aggregate data from Atlantic Information Services, the NCQA and the AHA to estimate the health plan demand model. I use a methodology similar to that first proposed by Berry, Levinsohn and Pakes (1995). The utility of individual i from plan j in market m is given by:

$$\tilde{u}_{ijm} = \xi_{jm} + z_{jm}\lambda + \gamma_1 EU_{ijm} + \gamma_2 \frac{prem_{jm}}{y_i} + \omega_{ijm}$$

where z_{jm} and ξ_{jm} are observed and unobserved plan characteristics respectively, $prem_{jm}$ are plan premiums, y_i is the income of individual i, and ω_{ijm} represents idiosyncratic shocks to consumer tastes, again assumed to be iid Type 1 extreme value. I consider HMO and POS plans only. The characteristics included in z are premium, the size of the physician network, plan age, a list of eight clinical quality variables (taken from the NCQA's HEDIS dataset) and two variables summarizing consumer assessment of plans on dimensions such as availability of needed care and speed with which care is received (from their CAHPS dataset). The results of this stage of the analysis are reproduced in Table 2. I find that consumers place a positive and significant weight on their expected utility from the hospital network when choosing a plan. The coefficient magnitudes imply that a one standard deviation increase in expected utility is equivalent to a reduction in premium of $39 per member per month (a little less than one standard deviation).

[14] It would be preferable to estimate consumers' hospital choices using data for managed care enrollees. However, this is not feasible because the available data do not identify the hospital networks offered by each managed care plan, so the choice set of managed care enrollees is unobserved. Instead I consider the choices made by indemnity and PPO enrollees, whose choice set is unrestricted. I assume that indemnity/PPO enrollees have the same preferences over hospitals as HMO/POS enrollees, conditional on their diagnosis, income and location. I test this assumption using data for HMO/POS enrollees in Boston; see Ho (2004) for details.

4.2 Producer Surplus Generated by the Network

With the demand estimates in hand, I now move on to consider the observed health plan-hospital contracts. The simplest model of contracting assumes that each insurer-provider pair bargains independently over the division of a surplus of size M. The implication (whatever the bargaining framework used) is that firms reach agreement if and only if $M > 0$. I investigate this theory by using my demand estimates to predict the producer surplus generated by each plan when it contracts with each potential hospital network: that is, the total profit to be divided between the plan and all the hospitals with which it contracts. The producer surplus generated by plan j in market m when it contracts with hospital network H_j is:

$$S_{jm}(H_j, H_{-j}) = \sum_i \left(n_i s_{ijm}(H_j, H_{-j}) \left[prem_{jm} - p_i \sum_{h \epsilon H_j} s_{ih}(H_j) cost_h \right] \right) \quad (1)$$

where n_i is the population in consumer-type cell i (defined by ZCTA, age and gender), p_i is the probability that a type-i person will be admitted to hospital, $cost_h$ is the average cost of treatment at hospital h, and $prem_{jm}$ is plan j's premium in market m. The quantities $s_{ijm}(H_j, H_{-j})$ and $s_{ih}(H_j)$ are plan j's and hospital h's predicted shares of type-i people when networks H_j and H_{-j} are offered by plan j and other plans respectively. These are predicted using the demand estimates and take account of the flow of consumers across plans, and across hospitals given their choice of plans, in response to network changes.

The surplus definition does not include plans' non-hospital variable costs. Each plan faces a number of costs of enrolling consumers: these include payments to primary care physicians and prescription drug costs, for example, in addition to the costs of treatment at hospitals. If these non-hospital variable costs differ across consumer types they may affect the plan's network choice since adding a new hospital could disproportionately attract certain types of consumer. Unfortunately, I do not have access to data on plan variable costs and therefore cannot include these effects in the surplus term[15]. However, I account for this issue later in the analysis by estimating the cost of enrolling certain types of consumers directly. The details of this robustness test are discussed in Section 7.

The calculation takes account of hospital capacity constraints. If any network combination implies that any hospital is over 85 per cent of its maximum capacity level, I reallocate patients randomly to non-capacity constrained hospitals in the market. I assume that patients are treated in the order in which they arrive and that the timing of sickness is random: each plan therefore has the same percentage

[15] The analysis does allow for the existence of additional fixed costs, since these would cancel out when we consider the surplus change from a change in networks.

of enrollees reallocated for any given capacity-constrained hospital. The adjustment affects patients' hospital choices and therefore their predicted costs of care but does not impact consumers' choices of plan or premium levels[16].

Premiums are assumed fixed in this calculation. In reality, when plan j considers a deviation from its observed network, it probably predicts that its own premium and those of other plans will adjust in response to the network change. Here again I encounter data limitations. I cannot estimate these adjustments accurately since I do not have access to panel data and so cannot observe the reaction of plan premiums to network changes over time. However, I include a robustness test for the fixed premium assumption; this is discussed in Section 8.

4.3 Does the Producer Surplus Term Explain the Observed Contracts?

The next step is to use the producer surplus estimates to identify the surplus generated by reversing each observed contract decision. I repeat this calculation for each of the 2706 potential contracts between the 451 plans in the data and the six largest hospitals in each market, keeping all other plans' networks fixed. The results are summarized in Table 3. I find that, for 54 per cent of the 2306 agreed contracts, the estimated surplus increase (which I denote $\Delta Surplus$) is greater than zero. For 52 per cent of the 400 potential contracts that were not agreed, the surplus that would be created by the contract is less than zero. So the simplest hypothesis explains the data in just over 50 per cent of cases. One way to interpret the fit of this simple model is to calculate a pseudo-R^2 measure. If we place equal weight on correctly predicting observed contracts and those that are unobserved, the pseudo-R^2 is 0.46[17].

I also conduct a probit analysis of an indicator variable for disagreement on the $\Delta Surplus$ measure. If producer surplus predicts contracts we should observe a negative relationship between these variables. However, I find no significant relationship when I include just these two simple measures in the regression. I then add an interaction between the capacity level in the market (measured as beds per thousand population) and surplus; that is, I estimate the following probit regression:

[16] A hospital is predicted to be over 85% of maximum capacity if predicted admissions * average length of stay at the hospital is greater than 85% of the number of beds * 365 days. By using the surplus variable without adjusting consumers' choices of plan, I am assuming that the plan does not expect consumers to predict their probability of treatment at each hospital in its network when choosing their insurer. Instead consumers are expected to assume they will have access to every hospital on the list. If a hospital is consistently capacity-constrained, consumers may update this belief. Unfortunately, without a panel dataset, there is no variation in the data to identify the extent of any such updating.

[17] The pseudo-R^2 is defined as $1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y}_i)^2}$, where y_i is the observed outcome, $\hat{y}_i$ is its predicted probability, and $\bar{y}_i$ is the mean value in the data. This is the same measure used to assess goodness of fit in Stata's logit and probit calculations. If an equal weight is placed on correctly predicting each observation, the model does much more poorly: in fact we find a negative pseudo-R^2.

$$\Pr(\textit{no contract}) = \Phi\left(\beta_0 + \beta_1 \Delta Surplus + \beta_2 capacity + \beta_3 \Delta Surplus * capacity\right)$$

where *capacity* is the number of beds per thousand population in the market[18]. The results are given in Figure 2. I estimate a positive value for β_1 and a negative β_3. Both coefficients are significant at p=0.1 (although the low pseudo-R^2 implies that the overall fit of this model is poor). Graph 1 in Figure 2 depicts the implied relationship between $\Delta Surplus$ and the probability of disagreement; graphs 2 and 3 divide this into a curve for high-capacity markets and one for low-capacity markets. In high-capacity areas the probability of disagreement is decreasing in the surplus measure, consistent with the simple theory. In low-capacity areas, however, the result is reversed: contracts that would generate a high producer surplus are the least likely to be observed[19].

5 The Price Negotiation

These results provide us with a puzzle. We have to explain the 46 per cent of contracts agreed when the predicted surplus increase is negative and the 48 per cent not agreed when $\Delta Surplus$ is positive. Further, we would like to understand why plan-hospital pairs in low-capacity markets should be particularly unlikely to reach agreement when they would generate a high producer surplus[20]. I therefore move on to consider the price negotiations which determine how the producer surplus is divided between insurers and hospitals. The following simple motivating example provides details on Stage 1 of the full game set out in Section 2.2. It demonstrates that, for certain types of provider, the negotiation can lead to selective contracting even when each additional contract would create positive incremental surplus, and to agreement when the incremental surplus is negative. In this section I outline the intuition and predictions of the model; details of the solution for a simple case are given in Appendix A. I make some fairly restrictive assumptions about the details of the bargaining game in order to solve for the equilibrium. However, the main intuition holds for more general models: I relax many of these assumptions when I return to the main empirical estimation.

Consider a market in which hospitals offer services to health plans. Insured consumers receive two

[18] Beds per thousand population is a crude measure of hospital capacity; the full model also takes account of demographic information.

[19] High-capacity areas are defined here as markets with more than 2.9 beds per thousand population, the average value in the data. The results are robust to clustering the error terms by plan or by hospital.

[20] Three assumptions made in the surplus calculation may help rationalize the contracts: the assumptions of fixed premiums, plan non-hospital variable costs, and exogeneity of other plan networks. All three assumptions will be addressed in the full model: I account for endogeneity of plan choices and include robustness tests for the importance of plan costs and premiums.

types of service from their plan: acute care from the hospitals in the network and preventive services from the plan's primary care physicians (PCPs). I assume the following order of actions:

1. Each plan announces its preferred hospital network
2. Plans agree on prices with their chosen hospitals
3. Plans set premiums; consumers decide whether to enroll and visit hospitals

Each plan-hospital agreement specifies the price to be paid to the hospital per treatment: the hospital is then required to treat every enrollee from the plan who requests care, provided it has spare capacity. I assume that plans make simultaneous, private take it or leave it offers to hospitals[21]. All firms have complete information about other firms' characteristics. Each plan offers a single product and sets a single premium level. Every consumer expects to need treatment once in the following year.

Each plan predicts the surplus created by every potential network, conditioning on its beliefs regarding other plan actions[22]. Each also calculates the profits that would be earned by every hospital at equilibrium; these would reduce the share of the surplus captured by the plan. Plans therefore choose their networks to maximize:

$$\pi_{jm}(H_j, H_{-j}) = S_{jm}(H_j, H_{-j}) - hospital\ profits \tag{2}$$

where $\pi_{jm}(H_j, H_{-j})$ is the profit secured by plan j when it contracts with hospital network H_j, given the networks of other plans H_{-j}, and $S_{jm}(H_j, H_{-j})$ is the producer surplus generated by the network[23].

In the simplest case consumers are unwilling to switch plans in order to gain access to hospital h. The hospital therefore needs to contract with every plan to maximize its expected number of patients: if it is not expected to be full it will accept any price offer that covers its costs[24]. Plans capture 100 per cent of the surplus created and will include hospital h in their networks provided it generates positive producer surplus. The more interesting situation arises when consumers are willing to switch plans if necessary to ensure access to h. In this case the hospital may credibly threaten to turn down low price offers for three reasons:

1. If all or most consumers with a positive value for h would switch plans to access it, turning down a low price offer would not significantly reduce demand for the hospital's services. In addition,

[21] This assumption simplifies the proofs for the theory model. A Nash bargaining model would not have qualitatively different results. In reality plans and hospitals probably bargain over contracts, implying that both insurer and provider capture a positive share of the surplus in each negotiation. I allow for this in the model for estimation set out in Section 7.

[22] Each plan is assumed to have perfect information regarding hospital characteristics. It can therefore perfectly predict which hospitals will accept its offers.

[23] I ignore plans' non-hospital costs here, but account for them in the empirical estimation in Section 7.

[24] Small hospitals, which would be full if they contracted with a single plan, are included in the category of capacity-constrained providers discussed in point 2 below.

by turning down the low offer, the hospital can force those enrollees who are willing to follow it across plans to move to an insurer that offers a higher price for their treatment. This increases the provider's revenues from these "switchers". The hospital will optimally turn down the lower offer if the revenues lost from enrollees it cannot then access are outweighed by those gained from consumers who are willing to move across plans. This is most likely for hospitals that are very attractive to consumers: I describe them as "star" hospitals. I discuss in Section 7 the characteristics used to define these providers in the empirical estimation.

2. The second scenario is that a subset of consumers will switch plans to access h and that the hospital expects to be capacity-constrained: that is, it can fill its beds without treating all consumers who wish to access it. In this case the loss of patients from the low-priced plan has a smaller (perhaps even zero) impact on the hospital's total volume, making it more likely to refuse the lower offer. A second effect also arises here. I assume in the theory model that the timing of sickness is random and that hospitals with spare capacity cannot turn away patients. This implies that, if the capacity-constrained hospital accepts all price offers, some enrollees from the lower-priced plan will displace higher-price patients from other insurers. Hospital h therefore has an additional reason to accept only the highest price offers[25].

3. Finally, if a sufficiently large proportion of the hospitals in the market merge to form a single system, the combined organization may be very attractive to consumers. As in case 1., the proportion of consumers willing to switch plans to access the system may be very high, implying that each system member will optimally turn down low price offers.

In all three cases, insurers will be forced to compete for contracts with the provider: hospital h will therefore capture a positive share of the surplus. A given plan may choose to exclude the hospital, focusing instead on those consumers whose low valuation for h and higher valuation for its other services prevents them from switching, if other plans have a higher maximum willingness-to-pay (WTP) for the contract[26]. The model predicts, therefore, that selective equilibria may be observed, even when positive incremental producer surplus would be generated from additional contracts, when hospitals are extremely

[25] The feature that distinguishes capacity-constrained hospitals from star providers is that the former might optimally accept the lower price offer if they had sufficient beds to treat all "non-switching" consumers who wished to access them. The latter have enough consumers willing to follow them across plans that they turn down low offers even when their beds are not full.

[26] A given plan may have a lower WTP for the contract than other insurers for two reasons. First, it may have a better outside option than other plans due to variation in consumers' preferences for other plan characteristics. Second, its non-switching enrollees may have a lower valuation for h than those in other plans: if so, the selective contract concentrates high-valuation consumers in the plans that contract with h, so that a higher proportion of their WTP can be extracted in the form of premiums.

attractive; capacity-constrained; or have merged to form systems.

The existence of systems can also explain the contracts that are agreed despite a negative incremental surplus. I observe in the data that some plans contract with some but not all members of a hospital system, but this practice is infrequent. I rationalize this observation with the idea that, if a hospital system has significant market power (as in point 3.), it will optimally impose penalties on plans that contract with some but not all of its members. Even systems with little market power may choose a bundling strategy, charging relatively more for contracts with individual hospitals than for those with the entire organization, to maximize the surplus captured from each plan. In both cases plans may be deterred from cherry-picking from the members of a system.

The model with capacity constraints is considered formally for the simple two-plan, one-hospital case in Appendix A. I show, first, that a hospital's ability to turn down low price offers forces plans to bid for contracts. Second, I demonstrate how variation in consumers' underlying preferences can provide one plan with a lower WTP for the hospital than the other's and can therefore lead to a selective equilibrium. Finally, I consider welfare effects. I choose the capacity constraints example for the detailed model because it raises the question: how can it be inefficient for a plan to exclude a hospital that is full in equilibrium? I demonstrate that an inefficiency can result even when no capacity is wasted and the consumers with the highest value for hospital h are the ones treated. It is generated because consumers are forced to make suboptimal choices across health plans in order to gain access to the hospital. The resulting loss of consumer welfare, which may outweigh the gain derived when the highest-valuation patients are given preferential access to h, would be avoided if both plans contracted with it. The intuition is similar for the examples concerning systems and attractive hospitals.

The theory model would be much more complicated for markets that contained large numbers of asymmetric firms: analytic solutions may well not exist in these more realistic cases. In order to consider real-life markets I therefore turn to the data and an empirical estimation of the profits captured by specific types of hospitals.

6 Reduced Form Analyses and Identification

6.1 Reduced Form Analyses

The theory outlined in Section 5 offers one possible reason why hospital-plan pairs which would generate positive producer surplus may not reach agreement: a hospital that does not need contracts with all plans may charge a sufficiently high price that only those with the highest willingness-to-pay agree to the contract[27]. In addition, the contracts that are agreed despite a negative estimated surplus may be explained by the penalties for excluding hospitals imposed by hospital systems. The full econometric model presented in Section 7 tests these theories by estimating the incremental profits captured by capacity-constrained providers, star hospitals, and system members. In this section I conduct three straightforward analyses to demonstrate the sources of identification for the model.

The first analysis concerns hospital systems. Table 4 shows the results of a simple probit regression of the following form:

$$\Pr(contract) = \Phi\left(\beta_0 + \beta_1 system + \beta_2 samesysdrop\right)$$

where *system* is an indicator variable for membership of a hospital system and *samesysdrop* measures the effect of the contract on the number of hospitals for which same-system members have been excluded by the plan. I also include market fixed effects. If the second system variable (*samesysdrop*) is excluded the estimated coefficient β_1 is positive, indicating that system members are on average more likely to agree contracts than other hospitals. When I add the second variable both coefficients become negative and significant. This implies that plans are less likely to agree to contracts with system hospitals than with other providers, and that they tend to contract either with an entire system or with none of its members.

The probit analyses reported in Table 5 are similar. Here I consider capacity-constrained hospitals: I use indicator variables for hospitals that were over 100 per cent and over 85 per cent of their maximum capacity in the previous year. The estimated effect of these measures on the probability of agreement is negative and significant, consistent with the theories set out above[28]. The analysis reported in Figure 2 and discussed in Section 4.2 also supports my theories regarding capacity-constrained hospitals. The finding that the probability of disagreement is increasing in producer surplus in low-capacity markets can now be rationalized by noting that, not surprisingly, capacity-constrained providers are most often observed in these areas[29]. It therefore makes sense that, in these markets, hospitals that offer a high

[27] The intuition is similar to that of a monopolist which restricts volume in order to maximize profits. Perfect price discrimination across consumers is impossible because enrollees are aggregated into plans, each of which charges a single premium, and may choose to move between plans. One effect of selective contracting is to concentrate high-valuation consumers into the plans that offer the highest prices to the hospital, reducing the negative impact of restricted volume on hospital profits. The main distinction between the monopolist example and the insurer-hospital case is that prices here are set by bargaining. The ability of a hospital to turn down low price offers also leads to an increase in its bargaining power with plans.

[28] Providers above their maximum capacity are defined as those with admissions $>$ 365*number of beds/average length of stay. The results remain significant when I also add the change in surplus when the contract is agreed and the two systems variables used in the previous analysis. They are also robust to clustering the error terms by plan or by hospital.

[29] For example, 12% of hospitals in markets with less than the average bed capacity, and only 7% of hospitals in other markets, were capacity constrained in 2001; the correlation between capacity constraints and beds per thousand population is -0.48.

surplus increase and are therefore most likely to be full should tend to contract selectively[30].

The final reduced form analysis considers the overall relation of market-level capacity to the extent of selective contracting in the market. The clustering of capacity constrained hospitals in low capacity areas implies that the probability of disagreement should be high in these types of market[31]. In addition, plans are obviously more likely to exclude hospitals in markets where this would have a lower effect on consumer demand. There is evidence that demand effects are lowest in markets with many beds per population[32]: the second prediction is therefore that the probability of disagreement should rise with bed capacity once it reaches a point where few hospitals are expected to be full. Thus a U-shaped curve is predicted, with a high probability of disagreement in high- and low-capacity markets and a lower probability in intermediate markets. To test this prediction I conduct the following probit regression:

$$\Pr(\textit{no contract}) = \Phi\left(\beta_0 + \beta_1 \textit{capacity} + \beta_2 \textit{capacity}^2 + \beta_3 \textit{capacity}^3\right)$$

The results are reported in Figure 3 together with a graph that summarizes them. The points on the graph show actual data: the per cent of hospitals excluded in each market plotted against actual bed capacity in the market. The fitted curve is U-shaped, as predicted by the model set out above.

6.2 Identification

The reduced form analyses demonstrate the variation in the data that will be used to identify the full model. For example, the incremental profits captured by capacity-constrained hospitals will be identified using variation in the probability of agreement across capacity-constrained and non-capacity-constrained providers, both within and across markets. The basic intuition is that, since we observe capacity-constrained hospitals refusing to agree to contracts more frequently than other providers, they must demand higher profits than their competitors. The producer surplus generated by these hospitals when they agree on contracts with particular plans provides an upper bound on the profits they capture. The predicted producer surplus generated when they do not reach agreement offers a lower bound on their profits. We can estimate the average profit of capacity-constrained hospitals by taking averages over these observations. A similar intuition applies to other types of hospital.

[30] The data confirms that high-surplus hospitals are the most likely to be full: the correlation between capacity constraints and $\Delta Surplus$ is 0.31.

[31] Hospitals in systems are not significantly clustered in either high- or low-capacity markets.

[32] For example the correlation between the market's capacity level and the average increase in expected utility when each hospital is added to a network containing all other hospitals is roughly -0.2.

7 A Model for Estimation

7.1 The Plan Profit Equation

The reduced form results lend support to the theories discussed in Section 5. However, all the analyses so far have ignored the potential interaction between plans' network decisions. The next step is to estimate a model that considers the information available to each plan when it makes its choices, and therefore takes into account the possible endogeneity. In order to do so we need an equation to describe plan profits.

The profit of plan j is the surplus generated given its chosen network H_j minus its costs:

$$\pi_{jm}(H_j, H_{-j}) = S_{jm}(H_j, H_{-j}) - c\_hosp_{jm}(H_j, H_{-j}) - c\_nonhosp_{jm}(H_j, H_{-j}) \tag{3}$$

where $c\_hosp_{jm}(H_j, H_{-j})$ is the cost of the plan's contracts with hospitals (generated by hospital profits) and $c\_nonhosp_{jm}(H_j, H_{-j})$ represents its non-hospital costs.

I assume that each hospital receives a two-part payment: a fixed element and a per-patient markup, $fc_{j,h}(.)$ and $mk_{j,h}(.)$ respectively. If prices are set by bargaining both these quantities depend on hospital and plan threat points and are therefore functions of characteristics of the hospital, the plan, and the market as a whole. I would ideally use a model of the plan-hospital bargaining process to estimate their values directly. However, the fact that each firm's threat point is endogenous (depending on the observed or expected outcome of all other pairs' negotiations), together with the number of insurers and providers bargaining in each market, makes this approach infeasible[33]. Instead I adopt a simpler methodology, closer to the reduced form analyses of the previous sections, by projecting hospital profits onto a set of hospital, insurer and market characteristics. That is, I estimate a reduced form function that summarizes the effect of these variables on hospital profits. This approach allows me to estimate the parameters of the model without specifying the exact form of the price negotiation. I relax several assumptions made in the model in Section 5: in particular I no longer assume that insurers make take it or leave it offers or that they have complete information about other plans' characteristics. The results from this relatively flexible model provide guidance for future modelling choices as well as information on the potential effects of policy and environmental changes on markups[34]. More specifically, the payment

[33] The model in Appendix A demonstrates that the contract between a given plan-hospital pair is affected by the outcome of the negotiations between all other pairs in the market: for example the WTP of other insurers and the availability and price demands of other hospitals all depend on the outcomes of their own negotiations and all affect the threat points of both plan and provider. Modelling this set of negotiations explicitly would be very complicated given that there are on average 12 plans and 15 insurers in each market.

[34] Prior attempts to analyze the fraction of the surplus that goes to hospitals look at the marginal value of the hospital

to each hospital is:

$$pmt_{jhm}(.) = fc(x_j, x_h, x_m)\vartheta_1 + N_{jhm}(H_j, H_{-j})mk(x_j, x_h, x_m)\vartheta_2 \tag{4}$$

where (x_j, x_h, x_m) are plan, hospital and market characteristics and $N_{jhm}(H_j, H_{-j})$ is the number of plan j's enrollees treated by hospital h[35]:

$$N_{jhm}(H_j, H_{-j}) = \sum_i n_i p_i s_{ijm}(H_j, H_{-j}) s_{ih}(H_j).$$

I allow for one source of randomness at this point[36]: an unobserved plan-specific fixed effect, $\alpha_j(.)$, that is constant across choices. This can be thought of as an "Aetna-type" effect that controls for the lower prices paid by plans with dominant market share. Including this term, and subtracting the sum across hospitals of the costs implied by equation 4, we obtain plan profits[37] as:

$$\begin{aligned}
\pi_{jm}(H_j, H_{-j}, x, \alpha, \vartheta) &= S_{jm}(H_j, H_{-j}) - \sum_{h \epsilon H_j} \left(fc(x_j, x_h, x_m)\vartheta_1 + N_{jhm}(H_j, H_{-j})mk(x_j, x_h, x_m)\vartheta_2\right) \\
&\quad - c_nonhosp_{jm} + \alpha_j(H_{-j}, x_j, x_m) \qquad (5) \\
&= \tilde{\pi}_{jm}(H_j, H_{-j}, x, \vartheta) + \alpha_j(H_{-j}, x_j, x_m) \qquad (6)
\end{aligned}$$

The third term in equation 5 relates to non-hospital costs. These are likely to vary with the sickness level of the individual patient. I would ideally estimate the cost of enrolling each type of consumer, defined by age and sex, by including the following expression in the profit equation:

$$c\_nonhosp_{jm}(H_j, H_{-j}) = \sum_i N_{jim}(H_j, H_{-j}) c_i$$

where $N_{jim}(H_j, H_{-j})$ is the predicted number of enrollees of type i in plan j given the equilibrium

to a network conditional on the networks in existence, but do not attempt to analyze the determinants of these networks. See for example Capps, Dranove and Satterthwaite (2004) and Town and Vistnes (2001).

[35] Note that both this number of patients and the surplus term also depend on the x's: the form of the dependence is modelled explicitly using the demand estimates from Ho (2004).

[36] A second error term, caused by observed differences in profitability that are not known to plans at the time decisions are made, is introduced in equation 7.

[37] A number of existing papers estimate the share of the surplus captured by a given hospital as a function of its characteristics and those of the plan and the market (conditional on the existence of the contracts; see Capps, Dranove and Satterthwaite (2003), Town & Vistnes (2001)). An analogous methodology would estimate the plan's share of the incremental surplus created when each hospital was added to the network. The current approach is similar. If we write $\pi_{jm} = \alpha(x_{h,} z_{jm}, \mu_{il}) S_{jm} = \sum_h N_{jh} \alpha_{jh}(prem_{jm} - \cos t_h)$, where α_{jh} is the share of the surplus retained by the plan when negotiating with hospital h, and ignore plan non-hospital costs, this is equivalent to: $\pi_{jm} = S_{jm} - \sum_h N_{jh}(1 - \alpha_{jh})(prem_{jm} - \cos t_{jh})$. The value estimated by the current methodology's markup term (if $fc_{j,h}(.) = 0$) is $mk_{j,h}(.) = (1 - \alpha_{jh})(prem_{jm} - \cos t_h)$. That is, in the absence of a prediction for the hospital's effect on premiums and therefore the surplus per patient, $mk_{j,h}(.)$ estimates not $(1 - \alpha_{jh})$ but the average profit per patient captured by the hospital.

hospital networks and c_i is the cost of insuring that type (to be estimated). Unfortunately the available data is not rich enough to estimate c_i in addition to the hospital cost parameters for more than one or two consumer types. In the main specification I set $c_i = 0$ for all i, assuming that non-hospital costs have little effect on plans' network choices. As a robustness test I consider the effect of including the predicted total number of plan enrollees and the fraction of enrollees aged 55-64[38]. The tests are discussed further in Section 8; they have little effect on the overall results.

The next step is to decide which variables to include in the expressions for fixed costs and markups. The list must be parsimonious: a large number of coefficients is unlikely to be identified given the limited data available and the fairly small variation in plan choice of networks observed. I use the theories discussed in Section 5 to inform the choice of variables. The main predictions are that hospitals that expect to be capacity-constrained, those in systems, and those for which all or most consumers would switch plans, should be most likely to fail to agree with plans (holding surplus fixed). System hospitals may also demand a higher price from a given plan if another same-system hospital is excluded than if it is not. Finally, even very simple bargaining models predict that lower-cost providers generate a higher total surplus, all else equal, and therefore earn higher markups than their competitors. I account for these predictions by including the following variables:

1. A measure of the extent to which particular hospitals are expected to be capacity-constrained. I derive an exogenous predictor of this variable by calculating the number of patients treated at each hospital under the thought experiment that every plan contracts with every hospital in the market[39].

2. Hospitals in systems, and those for which at least one same-system hospital is excluded.

3. Star hospitals: those that can expect enrollees to switch plans to ensure access. I identify these hospitals using the US News and World Report's hospital rankings for 2003, considering in particular each hospital's reputation ranking among physicians[40].

4. A measure of hospital costs per admission.

[38] I ignore the population aged over 64 because these consumers are eligible for Medicare.

[39] The hospitals that are "capacity-constrained" are those for which the predicted number of patients exceeds the number of beds * 365 / average length of stay in the hospital.

[40] US News magazine publishes an annual report giving hospital rankings for 17 different specialties and overall. The overall rankings comprise scores for reputation; severity-adjusted mortality ratios; and other care-related factors such as the number of nurses per bed and the technology available. The reputation score was compiled by asking a random sample of board-certified physicians which five hospitals they believed to be the best in their specialty. The rating is the percentage of responding physicians who cited the hospital. 583 of the hospitals in the sample (88%) have a reputation score of zero; the average nonzero value is 0.06 and the highest are 0.88 (Johns Hopkins Hospital, Baltimore) and 0.62 (Massachusetts General Hospital, Boston). I use the actual reputation rankings as a measure of star hospital status.

5. I also include a constant term in $mk_{j,h}(.)$: this identifies the average profit per patient received by non-system hospitals that are not capacity-constrained[41].

In reality the profit received by a particular provider depends not just on its own characteristics but on those of the plan and the market. For example, the price demanded by a system hospital can be no higher than the maximum willingness-to-pay of other plans in the market, and this depends on the attributes of other plans, consumers, and hospitals in the area. I would ideally include plan and market characteristics, and interactions with network attributes, to identify these effects. However, I have difficulty in identifying the coefficients on these terms[42]. It is perhaps unrealistic, given my limited data, to hope to estimate more than the most basic effects. The results reported therefore have no market characteristics: they identify only the average dollar profit per patient earned by each type of hospital[43].

There is not enough information in the data to allow for free interactions with both the fixed and the per patient component of the contracts. The results presented below are based on a specification where the fixed component of the contract depends on whether the hospital is in a system, whether another member of that system was excluded by the plan, and whether it was a star hospital, and the variable component depends on whether the hospital was capacity-constrained and the cost per admission of the hospital. When I estimated models allowing both sets of variables to affect both the marginal and fixed components, the individual coefficients were insignificant but there was little difference in the implications of the estimates.

7.2 Details on the Estimation Strategy

I adopt the methodology presented in Pakes, Porter, Ho and Ishii (2004) which uses a method of moments approach with inequality constraints. I use the necessary condition that plan j's expected profits from the observed H_j are higher than the expected profits from any alternative $H_{j,a}$. That is, I assume that:

$$E(\pi_{jm}(H_j, H_{-j}, x, \alpha, \vartheta)/I_{jm}) \geq E(\pi_{jm}(H_{j,a}, H_{-j}, x, \alpha, \vartheta)/I_{jm})$$

[41] This term enables us to move away from the strict take it or leave it offers model to a framework where all hospitals receive positive profits.

[42] Plan and market characteristics that are not interacted with network attributes do not vary across potential choices for a given plan and therefore cannot affect its choice. These characteristics therefore cannot be identified in the fixed cost term unless interacted with network attributes. It makes more sense to include these variables in the markup term, where they will be interacted with N_{jh}; however, in practice there was not enough variation in the data to generate significant coefficients.

[43] The capacity constraints variable is calculated using the predicted allocation of patients across hospitals when all plans offer a free choice: it therefore incorporates information on market characteristics. The other variables, however, relate only to hospital characteristics.

for all alternatives $H_{j,a}$ where I_{jm} is the information known to plan j when it makes its choice. Taking expectations here admits the possibility that each plan predicts its profits from any given network choice with error, perhaps because of inaccurate predictions of other insurers' equilibrium networks. I denote this error term ξ_{jH_j}, where:

$$\pi_{jm}(H_j, H_{-j}, x, \alpha, \vartheta) = E(\pi_{jm}(H_j, H_{-j}, x, \alpha, \vartheta)/I_{jm}) + \xi_{jH_j} \tag{7}$$

and:

$$E(\xi_{jH_j}/I_{jm}) = 0. \tag{8}$$

Now consider a finite subset of alternatives $H_j^A = \{H_{j,a})\}_{a=1}^A$. For any particular value of ϑ the value of $\pi_{jm}(H_j, H_{-j}, x, \alpha, \vartheta)$ is not observed (since the α is not observed). However, the vector

$$\Delta\pi_{jm}(H_j^A, H_j, H_{-j}, x, \vartheta) = \tilde{\pi}_{jm}(H_j, H_{-j}, x, \vartheta) - \tilde{\pi}_{jm}(H_{j,a}, H_{-j}, x, \vartheta)$$

where $\tilde{\pi}_{jm}(.)$ is defined by equation 6, is observable and can be computed. The assumptions imply that:

$$E\left[\Delta\pi_{jm}(H_j^A, H_j, H_{-j}, x, \vartheta) \ / \ z\right] \geq 0 \tag{9}$$

for all instruments $z \epsilon I_{jm}$. Note that the unobserved terms have dropped out of this equation. Equation 8 takes care of ξ, the error caused by the plan's inaccurate prediction of its profits from each network. The unobserved plan fixed effects are differenced out when we calculate the difference between the profits from observed and alternative choices[44]. Translating expectations into sample means, the equation for estimation is therefore:

$$\frac{1}{M}\sum_m \frac{1}{n_m}\sum_{j=1}^{n_m}\left[\Delta\pi_{jm}(H_j^A, H_j, H_{-j}, x, \vartheta) \otimes g(\ z)\right] \geq 0 \tag{10}$$

where M is the number of markets in the sample, n_m is the number of plans in market m, $\otimes$ is the Kronecker product operator and $g(\ z)$ is any positive-valued function of z. All ϑ that satisfy this system of inequalities are included in the set of feasible parameters. If no such ϑ exists we find values that

[44] If the plan had access to any information about hospital profits, or other characteristics that differed across networks, that was not available to the econometrician, these unobservables would not drop out of equation 9. The expectation operator conditions on the optimality of the observed choice as well as the plan's information set, so if the information (which I denote ν_j) affects the plan's choice, an additional term is generated of the form $E(\Delta v_j \ / \ z, \Delta\pi_j - \Delta v_j \geq 0)$.

minimize the sum of the absolute values of the amount by which each inequality is violated[45].

Identification in this model comes from comparing the profits of each plan when it chooses its observed network to those from its alternatives. For example, the identifying assumption implies that, if the plan is observed to contract with a capacity-constrained hospital, then the change in producer surplus it expects to result from the contract must be greater than the hospital's expected profits. If the plan drops the provider, the expected change in producer surplus must be less than those profits. Any feasible alternative networks could be used to generate these comparisons. I consider six alternatives $H_{j,a}$: these are defined by reversing the plan's contracts with each of the six largest hospitals in turn[46].

The instruments are required to be independent of the error term ξ_{jH}; they must also be positive (to ensure that no inequalities are reversed by the interaction with z). I use the characteristics included in the fixed cost and markup terms (the x's) other than the cost per admission, which I omit due to concerns about measurement error. I also include indicator variables for the following market and plan characteristics: a high number of beds per population, a high proportion of the hospitals in the market being in systems, a high proportion of the population aged 55-64, whether the plan is local, whether the plan has good breast cancer screening services and poor mental health services, and some of these characteristics interacted with the standard deviation of the distance between hospitals in the market (which as noted earlier is higher in selective than unselective markets because of its effect on consumer demand)[47]. None of these instruments is a function of the observed equilibrium. Each is known to the plan when it makes its choice. Each is also correlated with x: for example, plans can more easily exclude hospitals in markets with a younger, less sick population or with more beds per population. System hospitals are more often excluded in markets with a high proportion of hospitals in systems. The logic is similar for the other instruments.

The final step is to calculate confidence intervals for the parameter estimates. I use the simulation methodology described in detail in Pakes, Porter, Ho and Ishii (2004), estimating the limit distribution of the vectors used to define the inequalities, taking repeated draws on this distribution, and calculating a new estimate for each draw. The resulting vector of simulated values is used to find a 95 per cent

[45] One further assumption is needed. We have to assume that, when one plan deviates from its observed network, the others still succeed in securing the networks they bid for. This assumption seems reasonable since only small deviations are considered for just a single plan (j). In addition, the inequalities used for identification would still hold even if a plan was forced to drop one of the hospitals it chose to bid for, since such an action would increase the surplus generated by plan j. The parameter estimates should therefore be consistent even if the assumption is incorrect.

[46] The methodology could easily be extended to more alternatives per plan: for example, each could consider reversing two contracts at a time rather than just one. I try including these additional alternatives as a robustness test, and find little change in the overall results. I limit my main analysis to considering only six alternatives because of concern that the reduced-form function for hospital profits could change after a major network change.

[47] Low proportion means less than the mean percentile, except for beds per population and breast cancer screening rates where quartiles of the distribution were used.

confidence interval[48].

8 Results

8.1 Overall Results

The results are reported in Table 6. The estimate of ϑ for every specification was a singleton: that is, there was no parameter vector that satisfied all the inequality constraints[49]. The first column of the table reports results for the main specification. The point estimates all have the expected sign and have magnitudes which are consistent with the available information from other sources. Three of the five coefficients are significant at the traditional 5 per cent level, the "hospital in a system" indicator is significant at the 10 per cent level, and the constant term in the markup is significant at the 12 per cent level. However, the confidence intervals are reasonably large. The graphs in Figure 4 show the simulated distributions of three of the coefficients. They are clearly not Normally distributed. Each distribution is left-skewed, with most of its mass between zero and about a third of the upper bound to the confidence interval, but even with this caveat there is significant variance about the point estimates. This together with the robustness tests noted below makes statements about precise magnitudes difficult. The overall picture, however, is very clear. Hospitals in systems take a larger fraction of the surplus and also penalize plans that do not contract with all members. Capacity-constrained hospitals also capture high markups, and hospitals with higher costs per patient receive lower markups per patient than other providers.

To help interpret the magnitudes of the results, note that the average cost per admission for hospitals in the data is around $11,000[50]. The markups over these costs that I estimate vary by cost and type of hospital. For hospitals that are neither in a system nor capacity-constrained the point estimates imply very low average markups of around 2 per cent of revenues. Capacity constrained hospitals receive an extra $1800 per patient which, when their costs are taken into account, translates into an average markup of approximately 15 per cent of revenues. Hospitals that are not capacity constrained but are in systems capture $179,000 in incremental profits per month per plan, which given their average patient load translates into a markup of about $1800 per patient. When costs per admission are also taken into account, system hospitals are predicted to have average profits of around 14 per cent of revenues, 12

[48] The confidence intervals have not yet been adjusted to account for variance introduced by the estimated demand parameters. This is unlikely to significantly affect the results since the standard errors in this first stage were relatively low.

[49] As discussed in detail in Pakes, Porter, Ho and Ishii (2004), this does not imply that we should reject the specification. The result could easily be caused by the random disturbances in the markup terms.

[50] The cost variable, taken from the AHA survey 2001, is defined as total hospital expenses including items such as depreciation and interest expense.

percentage points higher than other hospitals. I also estimate a large penalty for excluding a hospital from a system: however, this happens only rarely.

The average hospital markup for community hospitals has been estimated at about 4 per cent of revenues[51]. This may be indicative that the estimated constant term in the markup equation is a little too small, particularly given that the markup from private payers is probably higher than that from Medicare and Medicaid, and that all payers are included in the published 4 per cent figure. The constant is the most imprecisely estimated of all the coefficients and if one looks at its distribution (shown in Figure 5) it is easy to see that the point estimate may well be lower than its actual value[52]. The other estimated coefficients, however, are not subject to this criticism.

The second column of Table 6 adds the US News reputation measure of star hospitals to the specification. This coefficient has a very large confidence interval, possibly because there are very few star providers in the dataset. However, adding this measure changes the other coefficients only slightly.

8.2 Robustness Tests

The results in the previous section support the theory outlined in Section 5: that hospitals in systems and those that are expected to be capacity-constrained seek rents and are optimally excluded by some plans in equilibrium. Hospitals in systems also seem to demand higher prices from plans that exclude their partners than from other plans. Could some other effect be causing these results? One possibility already discussed is that plans avoid contracting with hospitals that would attract enrollees for whom the non-hospital cost of insurance is high. The first robustness test therefore considers the effect of accounting for plan non-hospital variable costs. I include the number of enrollees and the fraction aged 55-64 in the plan profit equation; the results are reported in Table 7. The addition had little effect on the results: the key parameters actually increased in magnitude. The costs per enrollee are very imprecisely estimated but the magnitudes make sense: they imply a \$40 per month cost of insuring each young enrollee, and an additional cost of \$16,500 per month for every percentage point increase in the fraction of enrollees who are aged 55-64. This translates to an additional cost of approximately \$65 per month of insuring an older consumer.

Inaccuracies in the estimated demand system could cause problems with the capacity constraints variable. For example, if the demand for hospital h is biased up, so that the surplus increase when the

[51] The Kaiser Family Foundation report "Trends and Indicators in the Changing Health Care Marketplace, 2004 Update" provides data on hospital costs and profits. The average profit margin for community hospitals was 4.2% of revenues in 2001.

[52] If the true value was \$2900 per patient, rather than the estimated \$2600, this would imply markups of 4% of revenues.

hospital is added to plan j's network is inflated, this would also imply an upward bias on the estimated hospital profit[53]. This is an alternative explanation for the estimated profits of capacity-constrained hospitals: hospital h could be predicted to be full simply because its demand is biased up; both surplus and hospital profits would also then be (mechanically) overestimated[54]. I test for this by replacing the indicator variable for predicted capacity constraints with a variable less closely tied to the demand estimates: an indicator for hospitals that were full in the previous year. I choose not to include this variable in the main specification because it is endogenous. Any serial correlation in the disturbance from the model would induce a bias in its coefficient (and hospital-plan matches are quite stable over time). However, the endogeneity implies a negative bias, so a positive coefficient is still meaningful. The results are reported in Table 7. They are comparable to the main model: the capacity constraints coefficient is smaller than that in the main specification but still positive. This is reassuring; while the predicted surplus increase when hospitals are added to the network could still be biased up, it is not obvious why this should be more of a problem for hospitals that were capacity-constrained the previous year than for other hospitals.

The next test concerns the assumption of fixed premiums. If premiums would in reality fall when certain types of hospital were dropped then both the surplus increase from adding them, and the hospital profit needed to explain plans' unwillingness to agree contracts, would be biased down. I consider one robustness test for the fixed premium assumption. As already mentioned, I cannot account perfectly for premium adjustments in response to network changes because I do not observe premiums over time. However, I can use my estimated results to perform a robustness test. I allow all plans to simultaneously adjust premiums to maximize their profits (revenues less prices paid) where prices are determined by the estimates from the main specification. This premium adjustment is conducted as part of the surplus calculation for all networks considered; the supply side estimation is then repeated using the new measure of surplus. The results are reported in Table 7: they are fairly similar to those for the main analysis, although the magnitudes of some of the coefficients change.

Though none of the robustness tests change the qualitative nature of the results, some of the coefficients do change in magnitude. This is consistent with the results of a number of other robustness tests not reported here, and again implies that conclusions about overall effects can be drawn from the results

[53] This is particularly likely to be a problem if plan j is horizontally differentiated on a dimension not identified by the model. In that case the plan's estimated average quality would be biased up (to explain its ability to exclude hospitals); its increase in surplus when excluded hospitals are added would also be inflated. I exclude the plans that are most clearly horizontally differentiated from the dataset: these are Kaiser Permanente, Group Health Cooperative of Puget Sound and Scott & White Plan of Austin TX. It is possible that this is still an issue; however, the robustness tests described in this paragraph are reassuring.

[54] This argument would not, however, explain the reduced form results for capacity-constrained hospitals.

but that statements about precise magnitudes are difficult.

One further issue should be mentioned here: I have no data on plans' physician networks and therefore cannot account for them in the model. It is possible that plan j might decide not to contract with hospital h because this would involve establishing new physician contracts. There is no obvious reason why these physician contracting costs should be higher for a hospital that expects to be capacity-constrained than for other hospitals, but this point might go some way to explaining the result for hospital systems. If the physician networks associated with two hospital systems do not overlap, this provides an additional incentive for a plan to contract with all of one system or all of another rather than taking some hospitals from each. In the absence of relevant data it is difficult to say more on this issue; it may mean that the monetary costs of excluding a same-system hospital or contracting with a system hospital are overstated.

9 Implications of the Results

9.1 Explaining the Observed Contracts

The surplus term considered alone explained just over 50 per cent of the observed contracts. We can now compare the performance of the full model to this benchmark. I use equation 5 to predict the change in plan profits when each hospital is excluded from each plan's network, holding other plan choices fixed. I find that the new profit equation explains 62 per cent of the observed contracts overall: 66 per cent of the 400 contracts that were not agreed and 61 per cent of the 2306 contracts that are observed in the data. This is a considerable improvement on the 52 per cent and 54 per cent explained by the surplus term alone. The pseudo-R^2 value increases from 0.46 to 0.67. The 12 percentage point improvement in predicting the contracts that are not observed can be broken down as follows: approximately 25% of the improvement is explained by capacity-constrained hospitals; 42% by hospital systems, and the remaining 33% by the constant and cost per admission terms.

When we aggregate to the market level the clustering of types of hospital in specific areas also explains much of the cross-market variation in the data. The reduced form results show that markets with low and high bed capacity are more selective than intermediate areas. As already mentioned this can be explained by the higher proportion of capacity-constrained hospitals in low-capacity markets and the lower demand effects of selective contracting in high-capacity areas. The estimates are also consistent with the interview evidence that the dominant influence on network decisions belongs to insurers in some markets and to providers in others: it makes sense that hospitals should dominate both in low-capacity markets and in areas where many hospitals have merged to form systems. (In Salt Lake city, for example, two systems own six of the nine largest hospitals; we would expect hospitals to have high leverage here.) Plan power should be high in high-capacity markets with few systems.

9.2 Investment Incentives for Capacity Constrained Hospitals

The benefit that hospitals derive from capacity constraints implies a potential disincentive to invest that may have negative welfare effects[55]. I consider this issue both in general and for three specific hospitals that are predicted to be capacity-constrained: St. Luke's Medical Center in Milwaukee WI, SW Texas Hospital in San Antonio TX, and South Austin Hospital in Austin TX. For each I calculate the change in consumer surplus, plan profits and hospital profits that the model predicts would occur if the capacity constraints were removed.

Of course the removal of capacity constraints could affect plans' network choices. The model in this paper cannot predict the new equilibrium outcome for contracts, since as mentioned I do not fully detail how a single configuration of hospital networks is chosen from the multiple potential equilibria. In addition the change could lead to investment by other hospitals, or to entry or exit. The model cannot predict such developments. I therefore do not attempt a full analysis of the new equilibrium; I limit myself to a simple outline of the impact of the change if plans' choices of networks and all other market characteristics were fixed. Even this requires an additional assumption: that the institutional changes do not affect the reduced form function used to describe hospital markups. While not entirely realistic this enables me to derive at least a rough estimate of magnitudes. Finally, I also assume fixed premiums throughout the calculation.

The consumer surplus calculation finds the dollar value of consumers' gain in utility when reallocation of patients from full hospitals to other, less-preferred providers is no longer necessary. The utility gain is defined as:

$$\Delta CS_m = \sum_j \sum_i n_i s_{ijm} \frac{\gamma_1}{\lambda_i}(EU_{ijm}^{nocapcon} - EU_{ijm}^{capcon})$$

where EU_{ijm}^{capcon} is the utility a consumer with perfect foresight would expect to receive from the hospital network given the probability of reallocation away from the full hospital, $EU_{ijm}^{nocapcon}$ is the expected utility for the network when the capacity constraints are removed, and γ_1 and λ_i are the coefficients on the expected utility from the network and premium in the plan demand equation respectively[56].

[55] In terms of the model discussed in Section 5: if the hospital increased its capacity so that it was no longer capacity constrained, it might optimally accept all plan offers since the revenues lost from non-switching enrollees when it refused to agree to a low offer could now outweigh the increase in revenues from those who were willing to switch. Its threat to turn down low price offers would no longer be credible, so plans would revert to offering a price that just covered its costs and expecting the offer to be accepted.

[56] I assume that consumers do not have perfect foresight: they choose their plan under the belief that they can access

The plan profit calculation uses the expression for profits given by equation 5. The profit change from the removal of capacity constraints is given by:

$$\Delta\pi_{jm}^{capcon} = \pi_{jm}^{nocapcon}(H_j, H_{-j}) - \pi_{jm}^{capcon}(H_j, H_{-j})$$

where $\pi_{jm}^{capcon}(H_j, H_{-j})$ is calculated using the parameter estimates in Table 5 and $\pi_{jm}^{nocapcon}(H_j, H_{-j})$ sets the values for capacity constrained hospitals to zero. I adjust $S_{jm}(H_j, H_{-j})$ and $N_{jhm}(H_j, H_{-j})$ for the reallocation of patients across hospitals when the capacity constraints are removed. The change in hospital profits is calculated similarly. The reported results for plan profits include all plans in the market; the hospital calculation includes just the capacity-constrained hospital.

The results are set out in Table 8. The first row gives the median effect of investment to remove all hospital capacity constraints; the effects for three specific hospitals are then listed separately. The results are most easily understood by considering the specific examples which are reasonably representative of the overall data. The first two hospitals considered, St. Luke's in Milwaukee and SW Texas Methodist Hospital in San Antonio, are large providers with significant high-tech services: each has over 700 beds. The model predicts that each needs to roughly double its size to remove the capacity constraint. In both cases the change would have a substantial positive effect on consumer surplus: increases of \$0.16 and \$0.36 per person per new bed per year respectively. These figures translate to benefits of \$232,000 and \$469,000 per additional bed per market per year. The effects on producer surplus are smaller. The loss in hospital profits from reduced market power outweighs the increased revenue from new patients and implies that both hospitals would lose money from the change (even if it involved zero investment): the losses are \$18,618 and \$25,780 per year per additional bed respectively. In the case of St. Luke's the change would also reduce plan profits: the hospital has higher costs than others in the market, which are borne by the plan. Plans in San Antonio would see a positive but small profit increase of \$19,000 per bed per year. The third example is somewhat different and is representative of a second type of hospital that the model predicts to be capacity-constrained. South Austin hospital is a smaller suburban provider with fewer high-tech services, just 182 beds, and lower costs of care than its competitors. In this case the consumer surplus estimates translate to a \$326,000 benefit to consumers per year per additional bed. The hospital's low costs imply that plan profits would increase by a much higher \$61,000 per year per new bed and hospital profits fall by just \$5,300 per year per bed if the capacity constraints were

any hospital on its list. The utility reduction from capacity constraints is caused when consumers realize a lower utility from the network than they expected. I multiply by $\frac{\gamma_1}{\lambda_i}$ to convert from utils into dollars. Note that, in all three specific examples, only one hospital in the market is predicted to be capacity-constrained.

removed.

Overall, then, the results have three implications. First, the benefit to consumers of removing hospital capacity constraints is large: a median benefit of over $330,000 per person per year for each new bed provided. The available data on the average cost of new hospital capacity implies a payback period of less than two years when the impact on consumers is taken into account[57]. Second, plan profits increase as a result of the change even if premiums are assumed fixed. Any premium increases would further increase the benefit to plans: for example, an increase of 10 cents per person per year for each new bed would generate a median total gain in plan revenues (across all plans in the market) of around $165,000 per new bed per year[58]. Finally, however, in most cases hospitals (the organizations that are actually required to make the investment) face negative incentives to invest in this new capacity.

10 Discussion and Conclusion

The analyses in this paper demonstrate some of the causes of the observed hospital-insurer networks at the firm level. Four factors are important: consumer demand for a choice of hospitals; hospital costs of care; the extent to which hospitals can credibly threaten to turn down low price offers and therefore push prices up; and the existence of hospital systems. Together these rationalise the majority of the observed contracts.

The full model for estimation sets few limits on the structure of the negotiation: I simply assume that firms bargain over a price that has both fixed and variable components and impose Nash equilibrium conditions to estimate the parameters. Previous papers have modelled the negotiation in more detail than the very simple framework used here but have not accounted for some of the hospital characteristics that this paper shows are important. The results given here can therefore help determine which bargaining models best describe the hospital-health plan price negotiation; that is an additional contribution of this paper.

The estimates also relate to a fairly substantial literature on the effectiveness of HMOs and POS plans in controlling costs. The original rationale for managed care was that the threat of selective contracting could be used as a lever to prevent hospitals demanding high prices. A number of recent papers have set out interview and other evidence suggesting that health plans' leverage has declined in recent years

[57] Discussions with the COOs of several hospitals in the markets considered imply a capital cost of approximately $350,000 per new bed, assuming that a new wing is needed to house the new capacity, and staffing costs of around $65,000 per bed per year.

[58] The benefit to each individual plan is much smaller, however: an average of $14,000 per new bed per year if premiums increased by 10 cents per person. This small benefit, which implies a more than 20 year payback period for plans, explains why the prospect of profitable investment does not in general lead to vertical mergers.

prompting them to move away from selective contracting towards offering more choice[59]. The major causes of the reduced leverage suggested by these papers are a rising consumer demand for choice and an extensive consolidation of hospitals resulting in increased provider market power. Capacity constraints are also mentioned as a source of hospital leverage. The evidence set out in my first paper (Ho 2004) supports the first hypothesis: consumers do have a significant preference for choice. If this has developed recently, in response to experience of the restrictions imposed by managed care, it explains some of the move away from selective contracts. The results of this paper are among the first to support the other two hypotheses. Without access to data on actual prices paid it is impossible to know whether the reduced-form function estimated here has changed as a result of plans' selective contracting: that is, whether high-priced hospitals would demand even more if no plans turned them down. However, I do show that hospitals in systems and those that expect to be full are the most often excluded when the surplus they generate is positive, consistent with the theory that they have the highest leverage. Further research would be useful, particularly in a setting where price data is available, to investigate these issues in more detail.

The final implication of this paper relates to welfare. In my previous work (Ho 2004) I estimated that a move from the observed hospital networks to an unselective outcome would lead to a total gain in consumer surplus of $1.04 billion per year and a loss to producers of just $0.80 million per year across the markets in the sample, assuming fixed premiums. This paper demonstrates another important inefficiency caused by the contractual process: the high profits captured by capacity-constrained hospitals imply an incentive to under-invest in capacity that translates to a significant loss to consumers. Plan incentives are somewhat better aligned with consumer preferences, but are less relevant given that hospitals, not insurers, are the organizations required to make the investment. While the analysis involves a number of assumptions that limit the accuracy with which these welfare effects can be measured, the distortion to provider incentives is clear. This subject merits attention from both policy-makers and researchers.

[59] See for example Lesser and Ginsburg (2001) and Lesser, Ginsburg and Davis (2003).

Appendix A: A Theory Model for the Price Negotiation

Consider a market in which one hospital, H1, offers services to two health plans, P1 and P2. Both plans offer preventive services through primary care physicians; those at P2 have higher quality than those at P1. The market contains three types of consumers. Half the population places a positive value on P2's preventive services and the other half does not; of the consumers who value P2, half have a high valuation for H1 and half a low valuation[60]. The three types of consumer are therefore defined as follows:

$2N$ consumers of type $C1$, with WTP W_1 for treatment at H1 and 0 for P2's preventive services

N consumers of type $C2$, with WTP W_2 for treatment at H1 and V for P2's preventive services

N consumers of type $C3$, with WTP W_3 for treatment at H1 and V for P2's preventive services

Assume that $W_2 \geq W_1 > W_3$ and that $W_3 + V > W_1$. H1 has $3N$ beds and zero costs of care. Firms play the three-stage game described in Section 5.1. I assume that the timing of sickness is random. Therefore, if H1 agrees contracts with both plans and all consumers are insured, each consumer has probability $\frac{3}{4}$ of being treated at H1.

Finding the Nash Equilibrium of the game involves deriving each plan's payoff in each of the four possible network combinations. The case where neither plan contracts with H1 is straightforward: $\pi_{P1} = 0$ and $\pi_{P2} = 2NV$ (since P2 charges $prem_2 = V$). If only P2 contracts with H1, the plan predicts that H1 will accept any price that covers its costs; P2 therefore pays H1 0 per treatment. Algebra implies that firm profits are $\pi_{P1} = 0$ and $\pi_{P2} = \max(3NW_1,\, 2N(W_3 + V),\, N(W_2 + V))$.

In the case where only P1 contracts with H1, P1 pays $p_1 = 0$ per treatment. For simplicity I assume parameter values such that P1 chooses to charge premium W_1 and P2 lets the $C2$ types switch to P1 rather than cutting its premiums in an attempt to keep them. (That is, I choose parameter values to rule out price wars.) This assumption requires, first, that if P1 sets $prem_1 = W_1$ (offering $C2$ types utility $W_2 - W_1$), P2 chooses not to cut premium to the point where $V - prem_2 = W_2 - W_1$. The first assumption is therefore that $NV \geq 2N(V + W_1 - W_2)$, that is that:

$$2(W_2 - W_1) \geq V \tag{11}$$

The second requirement is that, given P2's unaggressive reaction, P1 will choose to set premium W_1

[60] Extending the model to allow non-equal proportions of different consumer types does not affect the qualitative results: no new effects are generated. The assumption that consumer tastes for health plan services (separate from their hospital networks) differ across plans is intuitive. Women aged over 55, for example, are likely to have a higher valuation than other consumers for plans with high breast cancer screening rates. The demand estimation described in Section 4.1 is limited in this regard: much of the utility derived from the plan is assumed to be constant across consumers. However, cross-consumer variation is introduced by the coefficient on premium/income and by the idiosyncratic taste term ω_{ijm}.

and attract both $C1$ and $C2$ types, rather than setting $prem_1 = W_2$ and hoping to attract the $C2$ types only[61]. The assumption is therefore that:

$$3W_1 \geq W_2 \tag{12}$$

Given these assumptions, firm payoffs are $\pi_{P1} = 3NW_1$ and $\pi_{P2} = NV$.

Finally, consider the case where both plans make offers to H1. There are two scenarios under which H1 will choose to accept just one plan's offer, implying an exclusive outcome. These are:

1. P1 has a higher WTP than P2, and if it wins, will choose to set $prem_1 = W_1$ so that both $C1$ and $C2$ types enroll in P1. H1 can then be certain of filling its beds even if it contracts exclusively with P1: if P1 makes a slightly higher offer than P2, H1 will accept just the higher offer.

2. P2 has a higher WTP than P1, and if it wins, will set $prem_2$ at a level to attract all three consumer types. Again, H1 will be certain of filling its beds by contracting exclusively with P2, and will do so if P2 makes a slightly higher offer than P1.

I consider here the parameter values needed to generate the more interesting case 1 since this potentially leads to an inefficient outcome. (Under case 2, all three types of consumer become enrolled in P2, implying that none had to make a suboptimal choice of plan to gain access to H1.)

We can determine whether P1 will choose to charge W_1, given any possible price p_1 paid to H1, by the following argument. If P1 pays price p_1 per treatment for an exclusive contract, it will choose to charge $prem_1 = W_1$ and attract $C1$ and $C2$ types if its profits from this option are greater than those from its alternatives[62]. Algebra implies that the choice will be optimal provided $p_1 < \frac{1}{2}(3W_1 - W_2)$. So if $p_2 < p_1 < \frac{1}{2}(3W_1 - W_2)$ H1 will accept P1's bid and reject that from P2.

P2's maximum price can be calculated by considering its alternatives. If P1 wins an exclusive contract, P2 will charge $prem_2 = V$ and sell only to $C3$ types, earning $\pi_{P2} = NV$. Conversely, P2 will win an exclusive contract if it offers a price higher than P1's and chooses to set $prem_2 = \frac{3}{4}W_1$ (attracting all three consumer types)[63]. P2's maximum WTP for an exclusive contract can be found by equating profits under the win and lose scenarios and requiring a price such that P2 chooses $prem_2 = \frac{3}{4}W_1$: algebra implies a value no higher than $p_2 = \frac{1}{2}(3W_1 - W_2 - V)$. Finally, H1 may agree contracts with both plans[64]. In this case P1 would set $prem_1 = \frac{3}{4}W_1$, attracting just $C1$ types. P2 would set $prem_2 = \frac{3}{4}W_3 + V$,

[61] It will never try to attract $C3$ types since in that case the hospital will be capacity constrained; the best it could do would be to set $prem_1 = \frac{3}{4}W_3$ and hope to earn $3NW_3 < 3NW_1$.

[62] These are to charge W_2 and hope to attract just $C2$ types or to charge $\frac{3}{4}W_3$ and hope to attract all consumers.

[63] Its other options are to set $prem_2 = W_2 + V$ or $W_3 + V$ and attract just $C2$-types or just $C2$ and $C3$ types respectively.

[64] This will be the outcome if the two plans offer equal prices.

attract both $C2$ and $C3$ types, and earn profits $\frac{3}{2}N(W_3 - p) + 2NV$. Equating this with P2's profits under the lose scenario implies that P2's maximum WTP for a nonexclusive contract is $p_2 = W_3 + \frac{2}{3}V$[65].

P2's maximum WTP for a contract with H1 is therefore given by $\tilde{p}_2 = \max(\frac{1}{2}(3W_1 - W_2 - V), W_3 + \frac{2}{3}V)$. The selective outcome will result provided that: $\tilde{p}_2 < p_1 < \frac{1}{2}(3W_1 - W_2)$. This is possible if the following inequality holds:

$$W_3 + \frac{2}{3}V < \frac{1}{2}(3W_1 - W_2) \tag{13}$$

The term on the left of this inequality represents P2's WTP for a contract; the term on the right is P1's incentive to sell to $C1$ as well as $C2$ types if it wins an exclusive contract with H1. The intuition is that P1 will secure the exclusive agreement provided P2 is willing to give up the contract and H1 expects P1's enrollees to fill its beds.

If equation 13 holds the outcome is therefore that H1 agrees an exclusive contract with P1; $C1$ and $C2$ types enroll in P1 and $C3$ types enroll in P2. Plan profits are $\pi_{P1} = 3N(W_1 - \tilde{p}_2 - \varepsilon)$ and $\pi_{P2} = NV$.

Can we assume that equations 11-13 can hold for reasonable parameter values? Simple examples demonstrate that this is possible: for example parameter values $W_1 = 100$, $W_2 = 120$, $W_3 = 60$ and $V = 40$ satisfy all three equations[66]. In these cases plan profits under the different network combinations (ignoring ε terms) are given by:

	P2 without H1	P2 with H1
P1 without H1	$0; 2NV$	$0; \max(3NW_1,\ 2N(W_3 + V),\ N(W_2 + V))$
P1 with H1	$3NW_1;\ NV$	$\min(3N(W_1 - W_3) - 2NV,\ \frac{3}{2}N(W_2 - W_1 + V))\ ;\ NV$

The potential Nash equilibria are (H1,0) and (H1, H1). In both cases the equilibrium outcome is selective: P1 contracts with H1 but P2 does not[67].

Now consider the constrained efficient outcome. Given the institutional restriction that the hospital is committed, once it signs a contract, to treating patients in a random order, the social planner compares the following scenarios:

1. The selective equilibrium outcome, which results in social welfare $2NW_1 + NW_2 + NV$

[65] Neither plan will choose to start a price war under the nonexclusive contracts provided both plans pay H1 a price p satisfying: $p > \max(2W_3 - W_1, 2W_1 - W_3 - \frac{4}{3}V)$. This inequality is satisfied at $p = W_3 + \frac{2}{3}V$. It is also possible that P1 would choose to charge premium $\frac{3}{4}W_1 + V$ and attract just $C2$ types, leaving $C3$ types uninsured. However, this would be optimal only if the price paid to H1 was very high; this is not the case for the example given here.

[66] It is also easily confirmed that P1 earns positive profits in the (H1, H1) case for these parameter values.

[67] In the second case, both plans bid for the hospital but only P1's bid is accepted. Two types of equilibrium are possible for other parameter values. First, if W_2 is sufficiently high compared to W_1, P1 would choose to set premiums to attract solely $C2$ types if it won an exclusive contract. In that case the hospital would choose to contract with both plans. Second, if W_3 is high, P2 may have a higher WTP than P1, and may itself secure an exclusive contract.

2. The unselective outcome where both plans contract with H1, P2 enrolls $C2$ and $C3$ types and P1 sells just to $C1$ types. Social welfare in this case is $\frac{3}{4}(2NW_1 + NW_2 + NW_3) + 2NV$.

The unselective outcome is therefore the constrained social optimum provided:

$$4V + 3W_3 > 2W_1 + W_2 \tag{14}$$

If this condition holds (as it does for the simple example given above), the benefit gained by $C2$ types from enrollment in P2 outweighs the loss created when lower-value $C3$ types displace other consumers from hospital beds. The inefficiency created by the bargaining process exists despite the fact that no capacity is wasted and that the consumers with the highest value for H1 are the ones treated. It is generated because H1 chooses which price offers to accept without internalizing the $C2$ types' benefit from enrollment in P2. A social planner takes this benefit into account when choosing the constrained efficient outcome.

References

1. Andrews, D., Berry, S., and Jia, P. (2004), "Confidence Regions for Parameters in Discrete Games with Multiple Equilibria, with an Application to Discount Chain Store Locations", working paper.

2. Berry, S., Levinsohn, J., and Pakes, A. (1995), "Automobile Prices in Market Equilibrium", *Econometrica*, 63(4), 841-890.

3. Brooks, J., Dor, A., and Wong, H., (1996), "Hospital-insurer bargaining: An empirical investigation of appendectomy pricing", *Journal of Health Economics*, 16, 417-434.

4. Burns, L., and Pauly, M., (2002), "Integrated Delivery Networks: A Detour on the Road to Integrated Health Care?", Health Affairs, 21(4), 128-143.

5. Burns, L., and Thorpe, D., (2001), "Why provider-sponsored health plans don't work", *Healthcare Financial Management*; 12-16.

6. Capps, C., and Dranove, D., (2004), "Hospital Consolidation and Negotiated PPO Prices", *Health Affairs*; 23(2), 175-181.

7. Capps, C., Dranove, D. and Satterthwaite, M., (2003), "Competition and Market Power in Option Demand Markets", *RAND Journal of Economics*, 34(4).

8. Ciliberto, F., and Tamer, E., (2003), "Market Structure and Multiple Equilibria in Airline Markets", Princeton and North Carolina State University Working Paper.

9. Cutler, D., McClellan, M. and Newhouse, J., (2000), "How Does Managed Care Do It?", *RAND Journal of Economics*, 31(3), 526-548.

10. Eggleston, K., Norman, G., and Pepall, L., (2002), "Managed Health Care and Provider Integration: A Theory of Bilateral Market Power", Working Paper, Tufts Department of Economics.

11. Feldman, R. and Wholey, D., (2001), "Do HMOs Have Monopsony Power?", *International Journal of Health Care Finance and Economics*, 1, 7-22.

12. Gal-Or, E., (1997), "Exclusionary Equilibria in Health-Care Markets", *Journal of Economics and Management Strategy*, 6(1), 5-43.

13. Gal-Or, E., (1999), "Mergers and Exclusionary Practices in Health Care Markets", *Journal of Economics and Management Strategy*, 8(3), 315-350.

14. Ho, K. (2004), "The Welfare Effects of Restricted Hospital Choice in the US Medical Care Market", working paper.

15. Lesser, C. and Ginsberg, P., (2001), "Back to the Future? New Cost and Access Challenges Emerge", Center for Studying Health System Change Issue Brief No. 35.

16. Lesser, C., Ginsberg, P. and Devers, K., (2003), "The End of an Era: What Became of the "Managed Care Revolution" in 2001?", *Health Services Research*, 38(1), Part II, 337-355.

17. Manning, W., Newhouse, J., Duan, N., Keeler, E., Leibowitz, A., and Marquis, M., (1987), "Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment", *American Economic Review,* 77, 251-277.

18. Mays, G., Hurley, R., and Grossman, J., (2003), "An Empty Toolbox? Changes in Health Plans' Approaches for Managing Costs and Care", *Health Services Research*, 38(1), Part II, 375-393.

19. Miller, R. and Luft, H., (1997), "Does Managed Care Lead to Better or Worse Quality of Care?", *Health Affairs*, 16(5), 7-25.

20. Pakes, A., Porter, J., Ho, K. and Ishii, J., (2004), "The Method of Moments with Inequality Constraints", working paper.

21. Seim, K., (2001), "Spatial Differentiation and Market Structure: The Video Retail Industry", Yale University Dissertation.

22. Town, R. and Vistnes, G., (2001), "Hospital competition in HMO networks", *Journal of Health Economics*, 20, 733-753.

23. Vistnes, G., (2000), "Hospitals, Mergers, and Two-Stage Competition", *Antitrust Law Journal*, 67, 671-692.

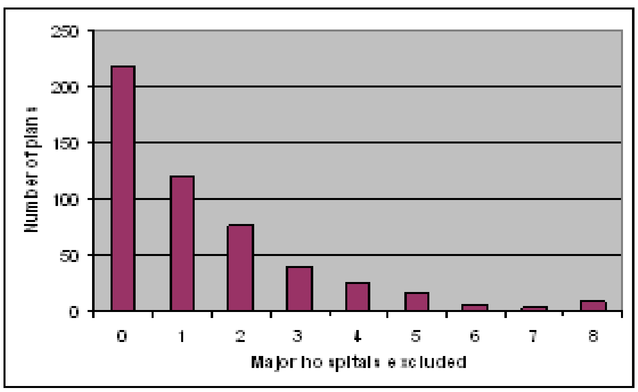
24. Zwanziger, J. and Mooney, C., (2000), "What Factors Influence the Prices Negotiated by HMOs and Hospitals?", Paper for the Seventh Northeast Regional Economics Research Symposium.

Figure 1: Variation in Plan Networks Across and Within Markets

This figure summarizes the variation in selectivity of plans' hospital networks both across and within markets. Markets are categorized on a scale from 1 to 5, where 1 is the least selective. Markets are fairly evenly distributed across the categories, as shown in the following table.

Category	Definition	Number of markets	Examples
1	The 5 largest plans (by enrollment) contract with all 8 largest hospitals (by number of admissions)	6	Baltimore MD; Atlanta GA
2	One plan excludes at least one hospital	10	Boston MA; Columbus OH
3	Two plans exclude at least one hospital or three plans exclude exactly one hospital each	6	Detroit MI; San Francisco CA
4	Three plans exclude at least one hospital; one of them excludes more than one	13	Houston TX; Miami FL
5	Four or more plans exclude at least one hospital each	8	Chicago IL; Los Angeles CA

Graph 1: Number of major hospitals excluded by each plan



Graph 2: Number of major hospitals excluded by each plan in selective markets (dark bars; categories 4-5 in the table above) compared to unselective markets (pale bars; categories 1-2 in the table)

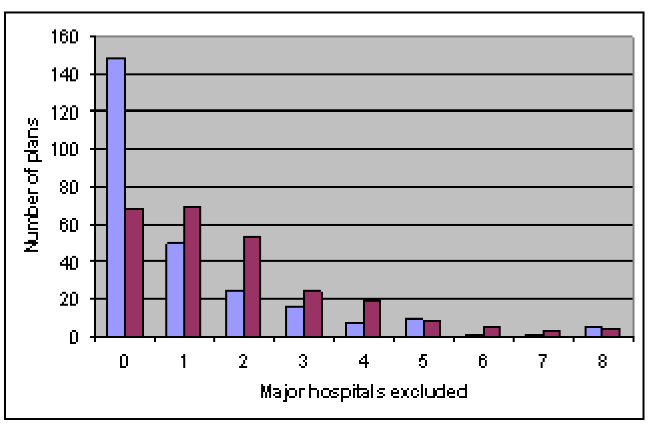


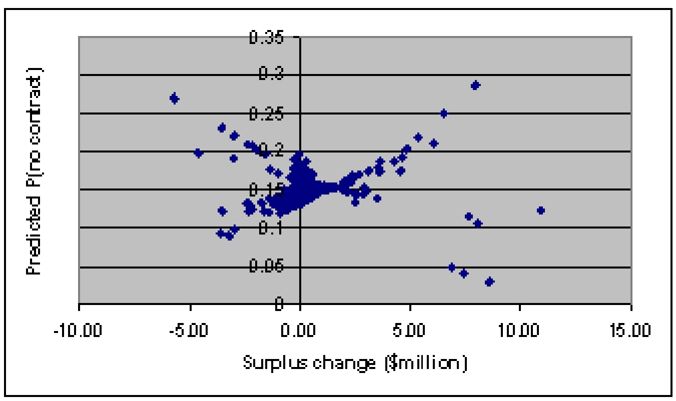
Figure 2: Effect of Surplus on Probability of Agreement

Probit analysis to predict disagreement using the change in surplus when a contract is agreed ($ million), bed capacity (measured in beds per thousand population), and an interaction term.

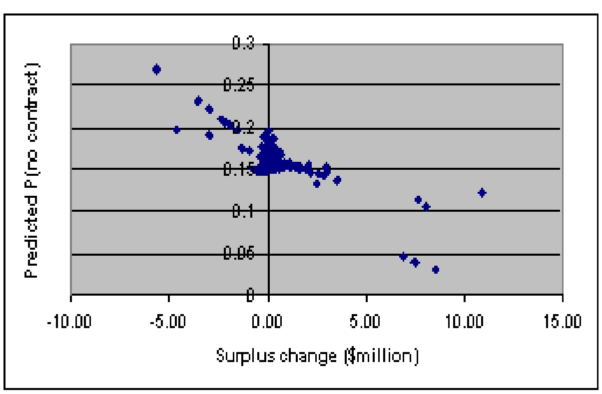
	Coefficient estimate
$\Delta Surplus$	0.152* (0.132)
Bed capacity	0.066* (0.037)
Capacity * $\Delta Surplus$	- 0.048* (0.041)
Constant	- 1.233** (0.108)
Pseudo-R^2	0.01

N=2706 contracts. Standard errors in parentheses; **significant at p=0.05; *significant at p=0.1

Graph 1: Effect of Surplus on Predicted Probability of Exclusion



Graph 2: High-Capacity Markets
(> 2.9 beds per 1000 popln)



Graph 3: Low-capacity Markets
(<2.9 beds per 1000 popln)

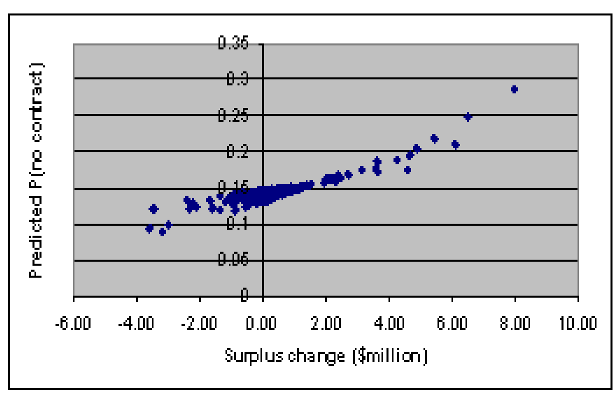


Figure 3: Effect of Hospital Capacity on Probability of Failure to Agree

Probit analysis to predict *disagree* using hospital bed capacity

	Coefficient estimate
Bed capacity (beds per 1000 population)	-5.86** (1.08)
$Capacity^2$	1.78** (0.33)
$Capacity^3$	-0.17** (0.03)
Constant	4.93** (1.12)
Pseudo-R^2	0.02
N	2706

** significant at p=0.05; * significant at p=0.1

Effect of Bed Capacity on per cent of Hospitals Excluded

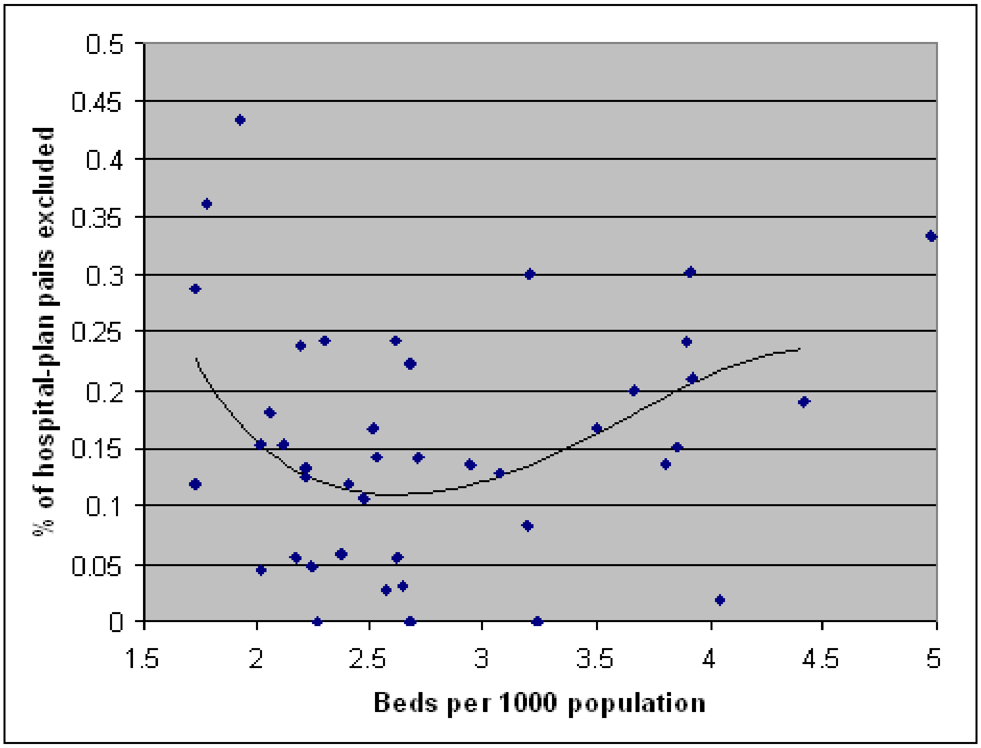


Figure 4: Simulated Distribution of Coefficients, Full Model

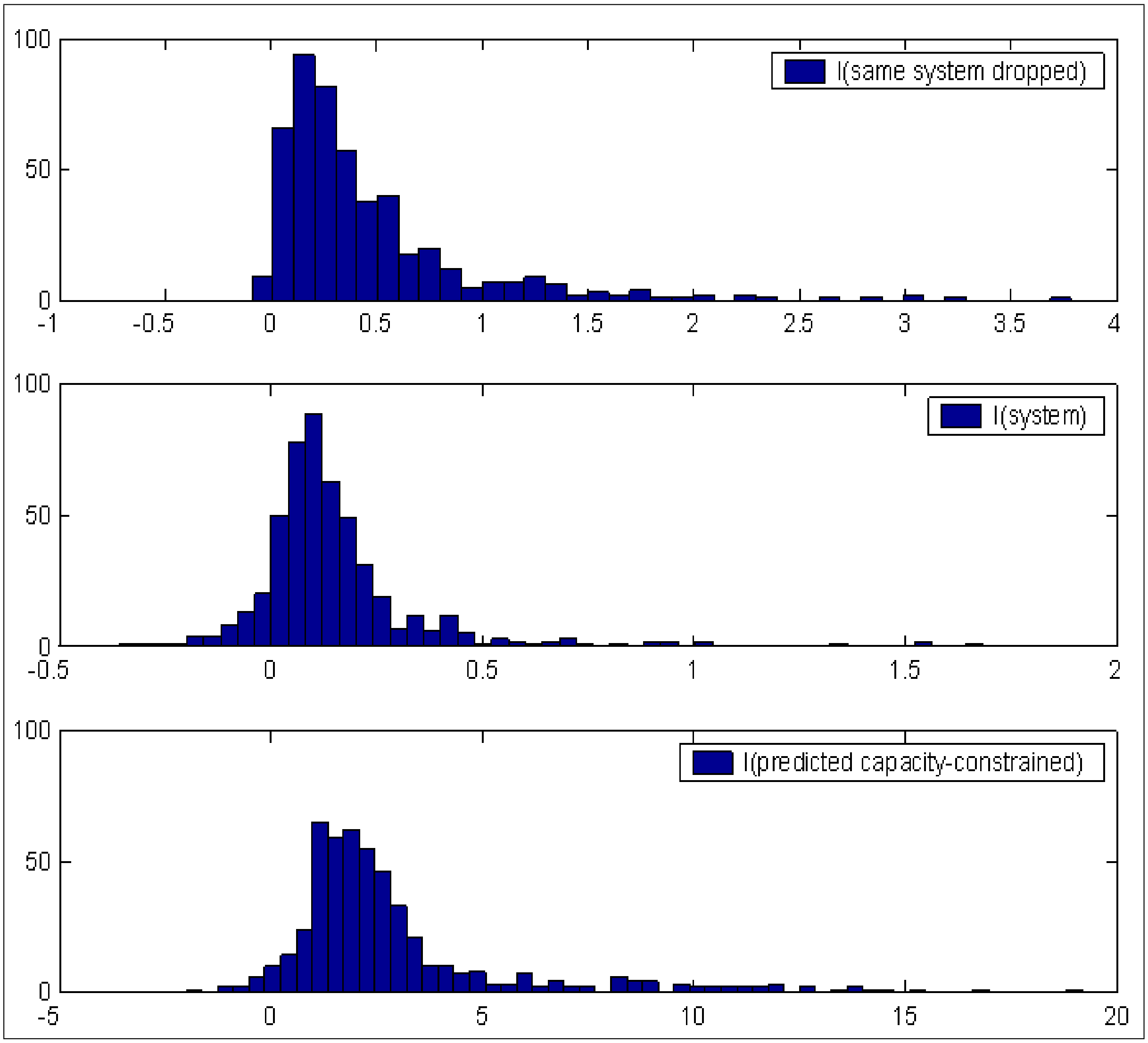


Figure 5: Simulated Distribution of Constant in Markup Equation

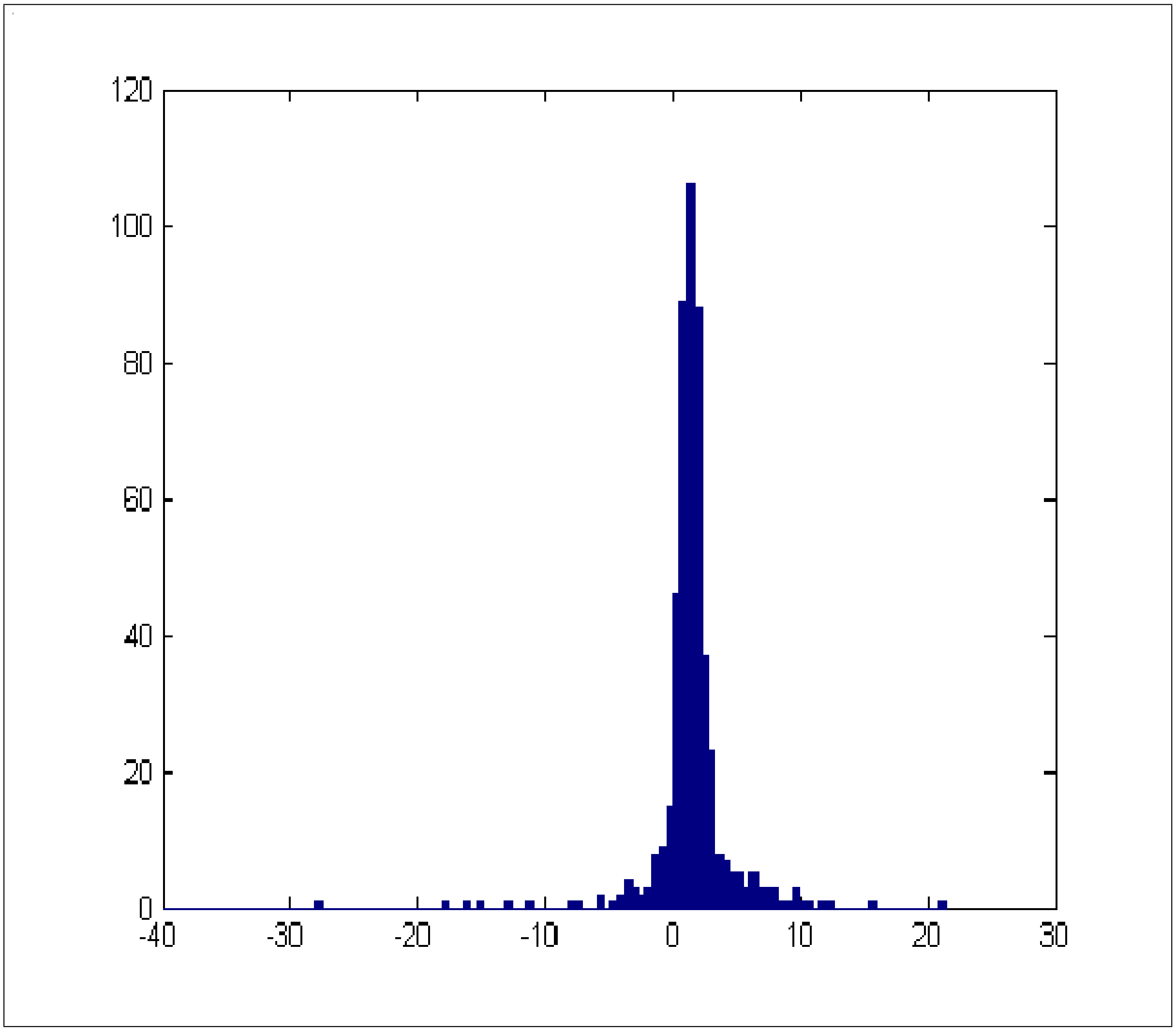


Table 1: Summary Data for Selective and Unselective Markets

	Unselective Markets (Category 1 and 2) Mean (std devn)	Selective Markets (Category 4 and 5) Mean (std devn)	p-value for difference in means
Market population (million)	2.34 (1.08)	2.36 (1.96)	0.95
Number of HMO/POS plans with over 1% market share	6.75 (1.65)	6.57 (1.89)	0.76
Number of hospitals	19.88 (10.99)	21.24 (20.53)	0.79
Beds per 1000 population	2.82 (0.96)	2.94 (0.96)	0.79
Managed care penetration	0.33 (0.16)	0.35 (0.15)	0.61
Average age of population	34.97 (2.28)	34.31 (1.39)	0.32
% of under-65 population aged 55-64	0.09 (0.01)	0.09 (0.01)	0.73
Median total family income of population	$49,160 ($8,244)	$46,130 ($8,642)	0.29
Mean distance between hospitals (miles)	11.62 (5.37)	12.54 (5.22)	0.31
Std devn of distances between hospitals (miles)	7.74** (3.26)	10.30** (4.06)	0.04
No. hospitals with open heart surgery	7.88 (3.63)	10.19 (8.59)	0.27
N	16	21	

** Means are significantly different at p = 0.05

Table 2: Results of Plan Demand Estimation, Full Model

	Coefficient Estimate
Premium ($00 pmpm)	-0.94 (1.13)
Expected utility from hospital network ($EUrep_{jm}$ or EU_{ijm})	0.59** (0.21)
Premium ($00 pmpm) / Income ($000 per year)	0.002 (43.9)
Physicians per 1000 population	0.21** (0.09)
Breast cancer screening	-0.38 (2.66)
Cervical cancer screening	4.40** (2.09)
Check-ups after delivery	0.18 (1.38)
Diabetic eye exams	-1.19 (1.60)
Adolescent immunization 1	-4.11** (1.17)
Adolescent immunization 2	3.08** (3.76)
Advice on smoking	6.17** (2.08)
Mental illness check-ups	2.70* (1.30)
Care quickly	0.78 (5.63)
Care needed	0.85 (3.99)
Plan age: 0 - 2 years	1.36* (0.97)
Plan age: 3 - 5 years	-0.64 (1.97)
Plan age: 6 - 9 years	-0.25 (0.58)
POS plan	-1.11** (0.13)
Constant	-10.50* (5.65)
Large plan fixed effects	Yes
Market fixed effects	Yes

N=559 plans. Standard errors (adjusted for the three-stage estimation process) are reported in parentheses. ** significant at p=0.05; * significant at p=0.1.

Table 3: Effect of Surplus on Probability of Agreement

The table reports the number of plan-hospital pairs for which surplus would increase, be unaffected, or decrease when the hospital was added, under the assumptions described in Section 4.2. For example, the data predicts an increase in surplus for 40.7 per cent of the 400 contracts that were not agreed.

	Number of contracts	$\Delta Surplus > 0$	$\Delta Surplus = 0$	$\Delta Surplus < 0$
Contract observed	2306	53.8%	2.2%	44.0%
Contract not observed	400	44.2%	3.8%	52.0%

Table 4: Effect of Systems on Probability of Agreement

Probit analysis to predict contracts using indicator variables for hospitals in systems. The second specification also includes the increase, when the contract is agreed, in the number of network hospitals for which a same-system hospital is excluded.

	Coefficient estimate	Coefficient estimate
System	0.079* (0.077)	-0.137* (0.080)
Same system hospital excluded		-0.769** (0.052)
Constant	0.466** (0.144)	0.495** (0.151)
Market FEs?	Yes	Yes
Pseudo-R^2	0.07	0.20

N=2706 contracts. Standard errors in parentheses; **significant at p=0.05; *significant at p=0.1

Table 5: Effects of Hospital Capacity Constraints

Probit analysis to predict contracts using hospital capacity constraints. Two measures are used: $cap\_constr1$ is an indicator variable for hospitals that were over capacity in the previous year (that is, those with admissions > 365 * number of beds / LOS in 2001); $cap\_constr2$ is an indicator variable for hospitals that were over 85 per cent of capacity in the previous year.

	Coefficient estimate	Coefficient estimate
$cap\_constr1$	-1.008** (0.234)	
$cap\_constr2$		-0.411** (0.110)
Constant	0.708** (0.145)	0.668** (0.142)
Market FEs?	Yes	Yes
Pseudo-R^2	0.08	0.08

N=2706 contracts. Standard errors in parentheses; **significant at p=0.05; *significant at p=0.1

Table 6: Results of Full Model for Estimation

This table reports the results of the full model. The coefficients represent the predicted profits to the hospital: a positive coefficient implies a positive hospital profit. "Hosp in system" refers to whether the hospital is in a system; "Drop Same System Hosp" refers to an indicator for hospitals for which a same-system hospital has been excluded; "US News Reptn" is the star hospital measure discussed in Section 7. Capacity constrained hospitals are those with predicted admissions (when all plans contract with all hospitals) > number of beds * 365 / average length of stay.

Charas of Hospitals	Coefficient Estimate	Simulated 95 per cent CI	Coefficient Estimate	Simulated 95 per cent CI
Fixed Component (Unit = $ million per month)				
Hosp in System	0.179	[-0.134, 0.624]	0.174	[-0.095, 0.543]
Drop Same System Hosp	0.595	[0.068, 1.810]	0.615	[0.057, 4.370]
US News Reptn			1.230	[-16.26, 2.579]
Per patient Component (Unit = $ thousand per patient)				
Constant	2.608	[-6.864, 11.21]	1.867	[-6.820, 9.550]
Capacity Constrained	1.867	[0.406, 11.39]	2.487	[0.130, 13.75]
Cost per Admission	-0.263	[-3.397, -0.20]	-0.131	[-2.471, -0.082]

N = 451 plans. 95 per cent confidence intervals in parentheses.

Table 7: Robustness Tests

This table reports the results of robustness tests for the full model. Test 1 includes the number of enrollees and the fraction aged 55-64. The second test replaces the predicted capacity constraints variable with an indicator variable for capacity constraints the previous year. Finally, the surplus estimate is adjusted for premium changes in response to changes in network. [NB: this test to be adjusted.]

Charas of Hospitals	Main Specification	Number Enrollees	Last Year Cap Con	Premium Adjustments
Enrollees		0.04		
		[-0.15, 0.16]		
per cent Aged 55-64		1.65		
		[-0.47, 5.49]		
Fixed Component (Unit = $ million per month)				
Hosp in System	0.179	0.50	0.28	0.59
	[-0.134, 0.624]	[0.06, 0.71]	[-0.24, 2.26]	[0.02, 0.80]
Drop Same System Hosp	0.595	0.68	0.85	0.57
	[0.068, 1.810]	[0.08, 2.10]	[0.07, 4.34]	[0.12, 2.12]
Per patient Component (Unit = $ thousand per patient)				
Constant	2.608	0.61	1.46	1.45
	[-6.864, 11.21]	[-0.68, 3.79]	[-9.83, 3.63]	[-4.03, 17.9]
Capacity Constrained	1.867	2.28		0.95
	[0.406, 11.39]	[-0.40, 2.79]		[-4.75, 14.7]
Last Year Cap-Constr			0.60	
			[-2.72, 6.71]	
Cost per Admission	-0.263	-2.09	-0.18	-1.67
	[-3.397, -0.20]	[-2.28, -0.31]	[-5.94, 0.04]	[-3.07, -0.30]

N = 451 plans. 95 per cent confidence intervals in parentheses.

Table 8: Investment Incentives for Capacity-Constrained Hospitals

This table shows the effect of investment to remove capacity constraints. The first row gives the median effect (across markets) of investment to remove all hospital capacity constraints. For some markets this involves investment in more than one hospital. Rows 2-4 list the effects for three specific hospitals: St. Luke's Medical Center in Milwaukee WI, SW Texas Methodist Hospital in San Antonio TX, and South Austin Hospital in Austin TX. All effects are given in $ per new bed per year. Plan profit effects are listed as a sum over plans in the market; hospital profits are given for the capacity-constrained hospital alone.

Example	CS per person ($ per bed per year)	CS per market ($ per bed per year)	Plan profit ($ per bed per year)	Hospital profit ($ per bed per year)
Median	$0.20	$338,800	$19,728	- $19,797
St. Luke's Milwaukee	$0.16	$232,270	- $2,779	- $18,618
SW Texas Methodist	$0.36	$469,000	$19,000	- $25,780
South Austin Hospital	$0.39	$326,400	$60,838	- $5,308