Issues in the Industrial Organization of Health Markets

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Kevin E. Pflum, B.S., M.A.

Graduate Program in Economics

The Ohio State University

2011

Dissertation Committee:

Dan Levin, Advisor
Paul Healy
Matthew Lewis
© Copyright by

Kevin E. Pflum

2011
Abstract

In the first chapter, a theoretical model of pharmaceutical sample dispensation is developed. We currently know very little about how samples are used and the reasons often presented in the medical literature do not have a theoretical foundation. The premise of the model developed in this chapter is that sample dispensation developed to mitigate the risk of trying a drug that, due to extreme idiosyncrasy in the consumer-to-drug matching, may have no therapeutic value to a consumer. Consumers are rational and in choosing, or not choosing, a drug so are maximizing the total, discounted expected utility of their choice. A firm’s sampling strategy depends on the degree of information available in the market. When the efficacy is commonly known the sampling decision is not monotonic in a drug’s efficacy. With asymmetric information, however, sample dispensation acts as a signal of efficacy and is monotonic. The signaling effect also causes the firm to dispense samples for lower efficacies than is optimal with symmetric information. The circumstances causing sampling to increase consumer welfare are explored.

The second chapter focuses on the problem of inducing socially preferred levels of quality when quality represents a dimension to a firm’s good or service that cannot be directly contracted upon. The existing mechanism design literature on regulation treats quantity and quality as interchangeable or observable. The inseparability
and non-contractibility of quality and quantity are explicitly modeled and show that
the optimal regulatory policy may generate a different outcome than has been pre-
viously suggested. Many markets possessing unverifiable quality often contain a mix
of for-profit and not-for profit firms; therefore, the firm’s objective is modeled as a
combination of profit- and output-maximization with a break-even constraint. The
firm’s ability to manipulate demand by adjusting quality can result in an under- or
over-supply relative to the first-best level, even for a pure profit-maximizer. A firm’s
informational advantage may be completely attenuated, however, when the firm is a
pure output-maximizer and the regulator removes the demand response to price by
paying directly for the good.

Finally, the third chapter examines the effect of hospital system membership on
the bargaining outcome between a hospital and a managed care organization (MCO).
Previous research has explored whether system hospitals secure higher reimbursement
rates by exploiting local market concentration—or market power—to increase their
value to MCO networks and negotiate a higher reimbursement rate. However, it is
also possible that system hospitals are able to leverage their system membership in
the bargaining game in some way in order to extract a higher percentage of their
value to an MCO network. We find that system hospitals that are highly concen-
trated within a patient market both generate a higher surplus from contracting with
a MCO and extract a higher proportion of the surplus. We also find that hospitals—
particularly for-profits—belonging to systems operating in multiple patient markets
have higher bargaining power suggesting that it is insufficient to only focus on local
patient markets when evaluating the potential effect of a merger. Our results indicate
that a majority of the price difference between system and non-system hospitals is attributable to differences in their bargaining power.
To Paula.
I owe a debt of gratitude to many people who have made this dissertation possible. First and foremost I must thank my fiancé, Paula Cordero Salas, who stood by me and encouraged me throughout the dissertation writing process. The process of getting a Ph.D. can push anybody to their limits and Paula has stood by me through the best and worst of the process. Without her love and support this dissertation may have never been completed.

Special thanks to my committee. First to my advisor, Professor Dan Levin, whose guidance and encouragement helped me to develop and focus my research. To Professor Matthew Lewis with whom I have spent many hours discussing this dissertation and from whom I have learned a tremendous amount. Last, but by no means least, to Professor Paul Healy who has gone above and beyond to provide guidance in all aspects of the Ph.D. process. Reflecting his generosity Professor Healy has provided me financial support so that can I focus on my dissertation, throughout which he has provided invaluable advice and support. Both inside and outside of the classroom Professor Healy is a role model whose example I hope to one day be able to live up to.

Many other people in the department deserve my gratitude. Professors Lixin Ye, Huanxing Yang, Jim Peck, and Yaron Azrieli have all provided valuable feedback and guidance at various stages of this dissertation. I am also grateful to numerous participants in seminars as the Ohio State University, the University of Michigan Ross School of Business, Kent State University, Bowling Green State University, Miami University, the University of Alabama, the Federal Trade Commission, and Charles
Rivers Associates as well as participants as the 2010 and 2011 International Industrial Organization conferences.

I would like to thank the Alumni association for providing me funds through the Alumni Grant for Scholarly Research to purchase some of the data used in this dissertation.

Finally, I wish to thank my family and friends. My parents, Tony and Nancy Pflum, have always supported by goals and believed in me. Their support forged my desire to achieve all that I could in life and I owe them everything.
Vita

1998 ................................. B.S. Mathematics, University of Victoria
1999-2000 ............................. Network Associate, Good Samaritan Hospital
2000-2005 ............................. Software Engineer, Strategic Analysis, Inc.
2005 ..................................... M.A. Economics, New York University
2006 ..................................... Senior Consultant, Booz Allen Hamilton
2006-2007 ............................. University Fellowship, The Ohio State University
2007-2009 ............................. Graduate Teaching Assistant, The Ohio State University
2009 ..................................... Department Citation for Excellence in Teaching, The Ohio State University
2009-2011 ............................. Graduate Research Assistant, The Ohio State University

Fields of Study

Major Field: Economics

Studies in:
   Industrial Organization
   Microeconomic Theory
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Dedication</td>
<td>v</td>
</tr>
<tr>
<td>Vita</td>
<td>viii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xiv</td>
</tr>
<tr>
<td>1. Utilizing Free Samples in Pharmaceutical Markets</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Background</td>
<td>9</td>
</tr>
<tr>
<td>1.2 The Model</td>
<td>13</td>
</tr>
<tr>
<td>1.3 Prices</td>
<td>19</td>
</tr>
<tr>
<td>1.3.1 Prices without Sampling</td>
<td>19</td>
</tr>
<tr>
<td>1.3.2 Prices with Sampling</td>
<td>21</td>
</tr>
<tr>
<td>1.4 Equilibrium Sampling</td>
<td>25</td>
</tr>
<tr>
<td>1.4.1 Symmetric Information</td>
<td>25</td>
</tr>
<tr>
<td>1.4.2 Asymmetric Information</td>
<td>28</td>
</tr>
<tr>
<td>1.5 Sampling with Informed and Uninformed Consumers</td>
<td>33</td>
</tr>
<tr>
<td>1.6 Welfare</td>
<td>37</td>
</tr>
<tr>
<td>1.6.1 Consumer Surplus</td>
<td>37</td>
</tr>
<tr>
<td>1.6.2 Socially Efficient Sampling</td>
<td>41</td>
</tr>
<tr>
<td>1.7 Discussion and Conclusion</td>
<td>42</td>
</tr>
</tbody>
</table>
## 2. Regulating a Monopolist with Unverifiable Quality 48

2.1 The General Model 58

2.1.1 Quality-Adjusted Cost and Value 61

2.2 Social Optimum 63

2.3 Regulating with a Marketed Good 65

2.3.1 Symmetric Information about $\theta$ 66

2.3.2 Asymmetric Information about $\theta$ 70

2.4 A Nonmarketed Good 84

2.4.1 Symmetric Information about $\theta$ 84

2.4.2 Asymmetric Information about $\theta$ 87

2.5 Asymmetric Information about Demand 90

2.6 Concluding Remarks 94

## 3. Diagnosing Hospital System Bargaining Power in Managed Care Networks 98

3.1 Background 103

3.1.1 Market Structure 103

3.1.2 Contract Negotiations 106

3.2 Related Literature 107

3.3 Model for Estimation 112

3.3.1 Willingness-to-Pay 112

3.3.2 Hospital Cost 115

3.3.3 Bargaining 117

3.4 Data 124

3.5 Results 128

3.5.1 Demand 128

3.5.2 Cost 131

3.5.3 Bargaining Power 134

3.5.4 Predicted Managed Care Patients and Bargaining Power 144

3.6 Market Power Versus Bargaining Power 145

3.7 Conclusions 147

## Appendices 166

A. Chapter 1 Proofs 166

B. Chapter 2 Proofs 173
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>10 Most Prescribed Therapeutic Categories, 2008</td>
<td>45</td>
</tr>
<tr>
<td>1.2</td>
<td>10 Most Sampled Prescription Drugs, 2008</td>
<td>46</td>
</tr>
<tr>
<td>1.3</td>
<td>Most Sampled Statins, 2008</td>
<td>47</td>
</tr>
<tr>
<td>3.1</td>
<td>All Hospital Characteristics</td>
<td>149</td>
</tr>
<tr>
<td>3.2</td>
<td>Private Hospital Operation Characteristics</td>
<td>149</td>
</tr>
<tr>
<td>3.3</td>
<td>Hospital Systems Operating in California</td>
<td>150</td>
</tr>
<tr>
<td>3.4</td>
<td>Top 10 Managed Care Companies</td>
<td>151</td>
</tr>
<tr>
<td>3.5</td>
<td>Discharges</td>
<td>153</td>
</tr>
<tr>
<td>3.6</td>
<td>Included Services by Diagnostic Category</td>
<td>154</td>
</tr>
<tr>
<td>3.7</td>
<td>Demand Estimates</td>
<td>155</td>
</tr>
<tr>
<td>3.8</td>
<td>Incremental Patient Value for Hospital System Membership</td>
<td>156</td>
</tr>
<tr>
<td>3.9</td>
<td>Cost Estimates: Variable Cost Characteristics</td>
<td>157</td>
</tr>
<tr>
<td>3.10</td>
<td>Cost Estimates: Fixed Cost Characteristics</td>
<td>158</td>
</tr>
<tr>
<td>3.11</td>
<td>Estimated Cost Differential for Not-For-Profit and For-Profit Hospitals</td>
<td>159</td>
</tr>
</tbody>
</table>
3.12 Estimated Cost Differential for System and Non-System Hospitals . . 160

3.13 Determinants of Bargaining Power . . . . . . . . . . . . . . . . . . . 161

3.14 System Characteristics and Bargaining Power . . . . . . . . . . . . 162

3.15 System Characteristics and Bargaining Power Adjusted for System
Market Power . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 163

3.16 Determinants of Bargaining Power and Patient Size . . . . . . . . . 164

3.17 Markup of Average Daily Reimbursement . . . . . . . . . . . . . . . 165
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Graphical Depiction of Proposition 10</td>
<td>80</td>
</tr>
<tr>
<td>3.1</td>
<td>California Managed Care Organizations Ordered by Market Share</td>
<td>152</td>
</tr>
</tbody>
</table>
Chapter 1: Utilizing Free Samples in Pharmaceutical Markets

In 1919 Mahon Kline of GlaxoSmithKline pioneered a new form of pharmaceutical marketing by mailing drug samples directly to physicians across the United States. As of 2006, the practice of “sampling” has expanded into a $2.6 billion marketing expense for pharmaceutical firms (Band, 2007).¹ The pharmaceutical industry is not alone in dispensing free samples either as samples can often be found with beauty and baby products, pet food, magazines and subscription television services. These markets all share the characteristic of consisting of nondurable experience goods (Nelson, 1974) that are subject to frequent repurchase. Idiosyncratic tastes in these markets make finding the most suitable product match to any particular consumer difficult for consumer and firm alike, especially in pharmaceutical markets, in which there is great idiosyncrasy in patient-to-drug matches. For example, a senior executive at GlaxoSmithKline made headlines when he revealed that 90 percent of drugs will work for between just 30 to 50 percent of patients!²

¹The value reflects the production cost of the samples and not the cost of delivery by representatives, on whom the industry spent $6.7 billion in 2006 (IMS Health, 2008). The retail value of the samples provided exceeds $16 billion and is the value often cited in the popular press.

The provision of free samples are generally welcomed by both physicians and consumers and pharmaceutical firms are eager to point out that samples help indigent patients get needed medication. The practice is not without controversy, however. For example the University of Michigan Health System has completely banned the practice and the University of Pennsylvania and Stanford University medical schools prohibit staff members from accepting samples (Rabin, 2007). Detractors often argue that the practice allows pharmaceutical firms to negatively influence a physician’s prescribing behavior (Backer et al., 2000; Balay-Karperien et al., 2007), and believe that there exist better alternatives to helping indigent patients such as voucher programs (Brennan et al., 2006; Greenland, 2009).³

The objective of this chapter is to develop a simple model of sample dispensation to provide insights into the circumstances leading to and consequences of dispensing free samples and to identify the welfare effects of sampling. As Arrow (1963) identifies in his seminal article on medical care, information asymmetries and uncertainty are critical features of health care markets to which its institutions have adapted. Consumers face considerable uncertainty concerning their idiosyncratic match with a drug, therefore the premise of this paper is that sample dispensation developed as a marketing tool that, by mitigating consumer risk, allows the firm to charge a higher price and increase market size from what they would otherwise be. We show that this role of samples along with the differences in the characteristics of therapeutic classes

³ Another marketing practice common in pharmaceutical markets that may be used to influence a physician’s prescribing behavior is the giving of small gifts, such as toys or pens, with the names of branded drugs as well as meals or conferences in idyllic locations. Konrad (2002) develops a model of promotional competition where gifting acts as a soft bribe entering the physician’s utility function.
and the knowledge of physicians and consumers generates sampling strategies that correspond well with what is observed in the data and the medical literature.

Our emphasis is on the role of the idiosyncrasy of consumer-to-drug matching in driving sample dispensation, therefore consumers are modeled as being forward-looking, rational economic agents who maximize their expected utility from treatment. Consumers’ foresight thus prevents firms from using samples to “hook” myopic consumers on their drug regardless of price and efficacy allowing for more nuance in the optimal sampling strategy. The firm has the standard objective of maximizing profits and does so by choosing its price and sampling strategy based on the market conditions and consumer beliefs.

We contrast sampling strategies from when there is symmetric information regarding the efficacy of a drug to when information is asymmetric and consumers have only a prior for the drug’s efficacy. When the efficacy of the drug is commonly known, the firm will provide samples to facilitate consumers in learning their ex post match value. The firm’s profit curve absent sampling (referred to as the no-sampling profit curve) is increasing and convex in the drug’s efficacy because efficacy affects profit in two ways. First, a higher efficacy results in a higher number of matches, thus a higher volume of repeat purchases; and second, the price the firm can charge is increasing in the efficacy as higher efficacy increases the option value of trying the drug.

In contrast, the firm’s sampling profit curve is less sensitive to the efficacy as a consequence of consumers learning their match before paying. Because the no-sampling profit curve is more sensitive to the efficacy than the sampling profit curve, when the two cross in efficacy space, they cross twice and the sampling strategy is
not monotonic in efficacy. Intuitively, there is no need to provide samples for a drug having an exceptionally high efficacy when the efficacy is known because there is little risk that the drug will not work for any given individual. In contrast providing samples for exceptionally low efficacy drugs is a losing strategy as well because few patients will find they are a match and return to purchase more of the drug.

The firm’s sampling strategy differs when information is asymmetric because consumers have two dimensions of uncertainty: uncertainty in the drug’s efficacy, and uncertainty in their idiosyncratic match. We show that the firm is unable to signal its efficacy through the price it sets; thus, consumers form a prior based on their knowledge of the efficacy distribution and the firm’s sampling choice. Because consumers use the firm’s sampling choice in forming their prior, firms may have to provide samples to signal their drug has a high efficacy, even when the firm would optimally not provide samples for a drug having a sufficiently high efficacy under symmetric information. The informational content of sample dispensation alters the firm’s strategy in two ways. First, it causes the firm to provide samples for drugs having a lower efficacy than it would under symmetric information. Second, the sampling strategy under asymmetric information is monotone in the drug’s efficacy as the firm will provide samples for all efficacies above some threshold. We characterize the market equilibrium, which may take the form of a pooling or a partially separating equilibrium depending on the values of the model’s parameters.

Finally we characterize the welfare effects of sampling. We show that consumer welfare increases with sample dispensation under one of two circumstances. First,
consumer welfare increases whenever the firm’s revenue decreases by permitting sampling. This outcome can occur only when information is asymmetric and consumers are sufficiently pessimistic because the firm’s revenue necessarily must increase if it optimally provides samples. Consumer welfare can increase when the firm’s revenue increases from sampling if the market-expanding effect of sample dispensation is sufficient to overcome the loss in consumer surplus that results from paying a higher price for the drug. If the firm is unable to alter the price consumers pay for the drug because of insurance, then sampling unambiguously increases consumer welfare. Otherwise sampling results in a net transfer of surplus from consumers to the firm. This result suggests that the least insured consumers are hurt the most by sampling. It follows that overall social welfare increases only if the market-expanding gain in consumer welfare exceeds the firm’s cost of providing samples.

Ours is the first game-theoretic model of product sampling of which we are aware. However, the model can be thought of as complementary to the analysis done by Joseph and Mantrala (2009) who look at how physicians utilize samples under diagnostic uncertainty. In their model physicians are able to diagnose a patient and determine which of two drugs is more appropriate before prescribing one; however, because of noisiness in the diagnosis, the physician is unable to determine with certainty which is the more appropriate drug. As a means of lowering the overall cost of treatment to the patient, the physician will provide the patient a sample and determine if they got the diagnosis correct before prescribing the drug. Joseph and

4In analyzing a signaling model of advertising Moraga-González (2000) treats sampling for products such as shampoo as a form of quality revealing advertising.
Mantrala focus on only the physicians’ strategy and not on the firms’ strategy, although they show samples give a pharmaceutical firm the ability to capture more of the market than they otherwise could. Choices of prices and the dispensation of samples are both taken as given making the model unsuitable for welfare analysis. Furthermore, although noisy diagnosis can explain sampling differences across therapeutic classes, it does not explain sampling differences across a drug’s lifecycle or the observed correlation between drug efficaciousness and sample dispensation. We similarly find that sample dispensation leads to more market coverage; however, the mechanism driving this result differs substantially. Moreover, we show that asymmetric knowledge of a drug’s efficacy and experiential learning can account for the characteristics of sampling that are empirically observed.

Although there has been surprisingly little work in the theoretical literature specifically on product sampling, the introductory pricing and quality signaling literature has some similarities to the asymmetric information regime of this chapter. Seminal articles include Shapiro (1983), Bagwell (1987), Liebeskind and Rumelt (1989) and Bagwell and Riordan (1992). A key distinction between the introductory pricing models and the current model is that the probability of a match is systematic in the introductory pricing models and idiosyncratic in the present model. The idiosyncratic match probability is the defining characteristic of pharmaceuticals which

5In the empirical literature, Crawford and Shum (2005) also consider idiosyncratic matches. In their model a patient receives normally distributed symptomatic and curative signals. Given the signals, physicians update their prior for the value of treatment with a given drug. Patients will switch treatments when a match is insufficiently low, analogous to a non-match outcome in the current model.
separate them from most other experience good markets, and as shown by Proposition 2, idiosyncrasy in the match outcome is the *raison d’être* for sampling. In the introductory pricing models the uncertainty is over a vertical quality characteristic and prices serve as a credible signal of the quality level because a mismatch between price and quality will result in a loss in demand as consumers learn their valuation for the good. That is, everybody is a match, but the value of the match varies from one consumer to the next. In the current model consumers are uncertain that they will be a match, and hence uncertain that the drug will have *any* value. Proposition 3 shows that the firm cannot use the price to signal the efficacy of the drug thus samples are used to mitigate this risk. Conditional on a positive match outcome, prices can be used to signal a vertical quality as in the introductory pricing and quality literature therefore to keep the focus on the role of samples to overcome the no-match outcome we fix the drug’s vertical quality as a static value.

Bergemann and Välimäki (2006) consider dynamic pricing of a new idiosyncratic experience good by a monopolist, which has some relation to the current model. They utilize a continuous-time model in which the information regarding a consumer’s value for the good arrives according to a Poisson process. Absent a signal, consumers balance the informational gain of continuing to purchase with the potential for current losses. The principal difference between their model and the current model is with respect to the nature of the uncertainty. In Bergemann and Välimäki (2006), the information regarding one’s match arrives according to a Poisson process. Thus, every consumer is essentially a match with the drug if they wait sufficiently long. More of the market becomes informed with the passage of time so the firm sets a price that
optimally balances extracting rents from the informed and keeping the uninformed in the market long enough to learn their ex post match value. A consumer’s decision to remain in or exit the market is based on the expected match value and the probability of receiving the signal. In the current model, the focus is on the idiosyncrasy of being a match and not on the idiosyncrasy of the ex post match value. Uncertainty in the match probability in the current model is similar to having uncertainty in the Poisson parameter of Bergemann and Välimäki.\(^6\)

Finally, there are several empirical papers exploring issues of pharmaceutical quality and learning (Berndt et al., 2003; Ching, 2010; Ching and Ishihara, 2010; Coscelli and Shum, 2004; Crawford and Shum, 2005; Mizik and Jacobson, 2004). Crawford and Shum (2005) examine physician and patient learning by estimating a dynamic matching model of demand uncertainty using data on anti-ulcer medication from the Italian pharmaceutical market and find substantial idiosyncrasy in efficacy across patients. Most important to the assumptions of the present model, Crawford and Shum find that consumers generally learn their match value to a drug very rapidly. Mizik and Jacobson (2004) seek to quantify the effects pharmaceutical detailing and free samples have on the number of prescriptions physicians write for the sampled drug. They find a modest, but statistically significant effect on the quantity of prescriptions written from both.

\(^6\)The relationship between the match probability of the current model and the Poisson parameter in Bergemann and Välimäki (2006) is not perfect; however, both have the same effect on a consumer’s option value of consumption today.
The remainder of the chapter is organized as follows. Section 3.1 provides some background on pharmaceutical sampling. The basic model is then introduced in Section 2.1. The firm’s pricing strategy given its sampling choice is derived in Section 1.3. The equilibria for both the symmetric and asymmetric information regimes is derived in Section 1.4. Section 1.5 examines the firm’s strategy when a portion of the market has symmetric information and knows the drug’s efficacy whereas the remaining portion of the market does not. Section 1.6 examines the welfare implications of sampling. Finally, Section 1.7 concludes with some remarks and suggestions for further research.

1.1 Background

To develop a picture of when samples are provided, Tables 1.1, 1.2, and 1.3 provide summary statistics for drug sampling in 2008 using data from the 2008 Medical Expenditure Panel Survey (MEPS). 7 Table 1.1 lists the top 10 most prescribed drug categories and reports the number of prescriptions written and the sampling frequencies for the category, the frequency of sampling for new prescriptions, and the frequency of sampling for refills. 8 The data show that there is considerable variation in sampling rates across therapeutic categories. For example, samples are provided for the eighth most prescribed category, topical agents, at a rate roughly three and a half times higher than the most prescribed drug category, cardiovascular agents.

7The statistics are restricted to drugs for which at least 35 unweighted prescriptions were observed in the data.

8Because MEPS only identifies whether samples were given to a patient in a particular survey round, the frequency of sampling for refills may be overstated.
Moreover, the data also show that samples are more likely to be provided to new prescribers.

In addition to reporting the sampling frequencies, Table 1.2 also reports the number of drugs with the same therapeutic class. Similar to the sampling rates at the category level, a sample is far more likely to be provided for a new prescription, though this varies from a high of about 13 times more likely with Nasocort AQ to just 1.2 times more likely with Yaz. All of the top 10 most sampled drugs are on-patent with at least two years left in the patent in 2008. Though at much lower rates, samples are provided for generic drugs as well.

Lastly, Table 1.3 reports the sampling frequencies for the popular drug category of statins as well as the number of patent years left for the statin in 2008. The distribution of samples between new prescriptions and renewed prescriptions are similar to the distributions at the category and drug levels for the newer, patented drugs, though there is no difference in sampling frequencies between new prescriptions and refills with the generic statins. Notably the drugs with more patent years left (which are also the new drugs) have the highest sampling rates. Lipitor is the most prescribed drug on the list (63 million prescriptions) followed by Simvastatin (47 million) and Crestor

\(^9\) Therapeutic classes are defined by Multum Lexico. The \# of drugs in a class is defined as the number of drugs belonging to the same Multum Lexico sub-sub-class thus represent the most likely substitutes.

\(^{10}\) For instance, the 10 least sampled drugs for which samples were given are all off-patent. Restricting the set to drugs which were prescribed at least 35 times in MEPS, the 10 least sampled drugs for which a sample was given are: Gabapentin (.0149), Cyclobenzaprine HCL (.0149), Lisinopril (.0145), Aspirin (.0144), Ibuprofen (.0125), Trazadone HCL (.0120), Acetaminophen/Codeine (.0117), Metoprolol Tartrate (.0110), APAP/Hydrocodone Bitartrate (.0095), and Hydrochlorothiazide (.0058).
Crestor is the most expensive and strongest of the statins whereas Lipitor is better than Simvastatin in some dimensions and Simvastatin is better in others. Venkataraman and Stremersch (2007) provides more evidence that sampling rates are correlated with the efficaciousness of a drug.

As profit-maximizing entities, it is clear that firms dispense samples to increase profits, and this most likely occurs through increased demand; however, there is no clear theory on how the availability of samples leads to increased drug usage in equilibrium (e.g., why don’t all drugs in a particular class provide similar numbers of samples?). There exist several theories, but these theories generally do not have a solid economic rational behind them and, in some cases, are inconsistent with empirical evidence. For example it is often argued in the medical literature that samples are used to simply persuade physicians to prescribe the sampled drug without concern for the optimality of the prescription choice. Related to this argument, some critics have claimed that samples are often targeted at more junior physicians because they are more susceptible to this form of influence (Rabin, 2007).

Boltri et al. (2002) is most frequently cited as evidence of these two arguments as the availability of samples was associated with physicians prescribing more costly ACE-inhibitors for hypertension over the recommendations of the Sixth Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure (JNC 6), which recommended diuretics and β-blockers be used as first-line treatments for hypertension. Though more expensive, the choice

---

11The unweighted prescription counts are 5,631, 4,269, and 1,937, respectively.

12Drugdigest.org provides detailed comparisons of statins and other drugs.
of ACE-inhibitors may have been optimal for the circumstance; however, as in 2003 the JNC 7 was released with new guidelines recommending that diuretics and ACE-inhibitors should be considered first-line treatments for hypertension. Additionally, the National Institute for Health and Clinical Excellence now recommends ACE-inhibitors alone as the first-line drug for hypertension for people under 55 years of age (National Institute for Health and Clinical Excellence, 2006). Thus, instead of making suboptimal choices the physicians surveyed by Boltri et al. may have been correctly optimizing patient treatments based on direct experience along with the availability of samples.

In addition to the study by Boltri et al. (2002), the evidence for the degree to which samples alter prescription choice are mixed as some studies indicate little change and others indicate significant change. For example, Brewer (1998) and Hartung et al. (2010) find very little difference in prescription choice when samples are not provided. However, Miller et al. (2008), Göñül et al. (2001), and Mizik and Jacobson (2004) find more significant increases in the number of prescriptions written for the sampled drugs for a variety of other therapeutic classes including nonsteroidal anti-inflammatories, oral diabetic agents, peptic ulcer and gastroesophageal reflux medications, and non-narcotic analgesics.

In summary, there is evidence that the availability of samples alter some physicians’ prescription choice in some therapeutic categories; but it is not clear how many explanations for these observations would exist in equilibrium. Given the variation in sampling rates across drugs, therapeutic categories, the point in a drug’s life-cycle,
and the timing of when samples are given to consumers, it is the premise of this chapter that the idiosyncratic matching, and uncertainty over the population-level match probability can account for these observations in equilibrium for an environment in which consumers act as rational, expected utility maximizers.

1.2 The Model

Consider a market where every period a continuum of mass 1 new consumers enter with the desire to receive treatment for a particular illness. Consumers can choose either a drug that is produced by one firm or an outside option reflecting either an alternative to medication such as diet or other lifestyle changes or another medication with a known match value; e.g., an older medication that has been on the market for an extended period of time. Consumers embody both the role of physician and patient. Given the potential importance for agency issues, this modeling choice requires some justification. Combining the roles is an issue only if the physicians have separate objectives from patients. Because of professional ethics and the hippocratic oath, however, physicians are often assumed to have an altruistic component to their objective function\textsuperscript{13} bringing the incentives of the physician and patient into closer alignment.\textsuperscript{14} Moreover, it is not clear that a physician has different financial incentives to prescribe one drug over another.\textsuperscript{15} Because the focus of our analysis is on how

\textsuperscript{13}See, for example Ellis and McGuire (1986), Ellis and McGuire (1990), and Chalkley and Malcolmson (1998b).

\textsuperscript{14}See Mott et al. (1998) for a discussion of the potential agency, or common-agency issues when the physician is limited by drug formularies.

\textsuperscript{15}Clearly a physician has a financial incentive to prescribe one drug over another when she is paid on behalf of a pharmaceutical firm to be a speaker or consultant, but this is a separate issue from sample dispensation.
samples may be strategically used to facilitate learning, augment beliefs, increase profit, and/or induce treatment, regardless of to which party these effects can be attributed, we feel the simplification is appropriate.\textsuperscript{16}

To keep the focus on the role of the sample for a product which may have no value to a consumer the efficacy $\phi \in [0, 1]$ of the drug is defined as a Bernoulli-distributed match probability which the firm is unable to select.\textsuperscript{17,18} When the drug’s efficacy is not commonly known by consumers their prior is represented by the distribution $G$ having strictly positive density $g$, which is known by the firm.

The marginal cost of production is independent of the efficacy level of the drug, so is normalized to 0. Dispensing samples has a cost of $c_S$ per period. The cost of providing samples is thus assumed to represent primarily a fixed cost consistent with the assumption that the production cost is relatively small so normalized to 0. In contrast, the cost of maintaining drug representatives and an infrastructure, from which the firm can dispense samples, is assumed to be relatively costly. We could include a variable cost to sampling, perhaps consisting of the opportunity cost of dispensing samples to consumers who would have taken the drug anyway or from physicians who take samples home with them, in addition to a fixed cost; however,

\textsuperscript{16}Examining what agency and behavioral issues affect a physician’s prescription choice remains an important avenue for future research and will require a thorough understanding of the behavioral issues relevant to treatment choice.

\textsuperscript{17}The advertising and introductory pricing literature addresses the signaling problem when the ex post match value is a random variable and there is no mass point at zero.

\textsuperscript{18}The efficacy of a drug is predominantly learned by the firm and the Federal Drug Administration (FDA) during the trials phase of development. For the FDA to approve a drug, among other factors, it must be demonstrably more effective than a placebo, therefore very low efficacy drugs are prevented from entering the market. For expositional simplicity, however, we ignore this restriction by allowing the realized efficacy to be as low as 0.
this modeling choice does not alter any of the qualitative results so for expositional clarity there is only a fixed cost to sampling.\textsuperscript{19} The firm’s profit, which is dependent on its sampling strategy, is defined in the next section.

Consumers are indexed by $i \in [0, 1]$. Consumer $i$’s monetary value of being treated is measured by $\theta_i \in [\underline{\theta}, \overline{\theta}]$ and is known by the consumer, but not the firm. To understand the meaning of $\theta$, consumers can be thought of as having a baseline healthy utility of $H$ and illness severity $\tau$. If we let $V$—where $V$ is a real-valued concave function—represent the consumer’s monetary value for health and the drug completely remedies the illness, then the value of treatment for patient $i$ is given as $V(H) - V(H - \tau_i) \equiv \theta_i$. We will refer to $\theta$ simply as the severity of illness. The firm knows the distribution of illnesses, $F$, and its properties. This mirrors the observation that pharmaceutical firms leave samples with physicians, so do not observe the illness severity of any given patient, but through market research are aware of the distribution of the illness within the population. The distribution $F$ is assumed to be continuously differentiable and has strictly positive density $f$ over the supports. To further ensure that the profit maximization problem is well behaved, $F$ satisfies the monotone hazard rate property $d\{F(\theta)/f(\theta)\}/d\theta > 0$ for all $\theta$ in the support.

Reflecting evidence that patients may learn their match value for a drug relatively quickly,\textsuperscript{20} a consumer learns if he is a match to the drug after the first course, or

\textsuperscript{19}What is important to derive the results is that sampling is costly to the firm as reflected by the billions of dollars spent by pharmaceutical firms each year on sample dispensation.

\textsuperscript{20}See for example Crawford and Shum (2005), who observe that patients learn their match value with anti-ulcer medicine “quite quickly” and generally after the first use.
period, of treatment.\textsuperscript{21} If the consumer is a match with the drug, then he continues the treatment by purchasing the drug for a total of $T$ treatment periods.\textsuperscript{22} At the conclusion of the treatment the consumer is cured and exits the market. Because a unit mass of new consumers enter and a unit mass of consumers exit the market every treatment period, the size of the market remains constant. Moreover, $T$ generations of consumers are in the market in any given treatment period; i.e., the market is in a steady-state.\textsuperscript{23}

The consumer-to-drug match value $m_i \in \{0, 1\}$ is 0 in the event of a non-match and 1 in the event of a match. Consumers’ period utility from treatment can be expressed as

$$U_{i,t} = \theta_i m_i - p_t \text{ for all } t \in \{1, \ldots, T\},$$

where $p_t \geq 0$ is the price set by the firm. In the presence of insurance, $p_t$ represents the price the consumer is responsible for when they have to pay a co-insurance rate. We also consider the firm’s strategic use of samples and the welfare effects of sampling when consumers are fully insured and are responsible for a co-pay that is not altered by the dispensing of samples. Prescription drug markets in the U.S. will generally contain patients with a mix of insurance and payment responsibilities, nevertheless

\textsuperscript{21}A period of treatment is thus defined as the length of time required to learn if a drug is working, e.g., a few days, a week, or a month, and not defined as a single dose.

\textsuperscript{22}The number of treatment periods need not be deterministic. For example, after consuming the drug for one unit of time, it may be the case that the consumer needs to continue the treatment with probability $\rho$, which is independent of the drug’s efficacy or characteristics but is known to the firm. When the consumer continues the treatment with probability $\rho$, then $T = \frac{1}{1-\rho}$ is the expected total number of treatment periods.

\textsuperscript{23}Factors that would alter the market’s steady-state include direct-to-consumer advertising because it has the potential to increase the mass of entering consumers and demographic changes such as a growing elderly population.
the cost of prescription drugs remains an important characteristic of the market as nearly half of the amount spent on prescription drugs is out of pocket compared to 20% for other healthcare expenditures (U.S. Department of Health and Human Services, Medical Expenditure Panel Survey, 2008).

Each consumer $i$ also has available an outside option $\bar{u}(\theta_i)$ where $\bar{u}$ is non-negative and increasing. Furthermore, $\bar{u}(\theta) \geq \theta$ and $d\{\theta - \bar{u}(\theta)\}/d\theta \geq 0$ for all $\theta \in \Theta$. These last two properties are regularity conditions establishing that higher types weakly value the drug more relative to lower types (e.g., because of risk-aversion) and lower types may prefer their outside option to the drug. Consequently, if there is a type $\tilde{\theta}$ such that all consumers of this type are just indifferent between trying the drug and the outside option, then all types $\theta > \tilde{\theta}$ will weakly prefer the drug. The assumption is that consumers have a prior for the value of their outside option and for treatment with the drug. In reality this expectation is formed with the help of physicians who diagnose patients identifying to some degree the patients’ severity of illness and type. Based on the diagnosis the physician and patient then pick the treatment option optimizing the present discounted utility of the patient. Finally, although consumers’

---

24 Even with Medicare Part D, the prescription drug-coverage component of Medicare, consumers may have very high out-of-pocket costs. For example, a beneficiary pays 100 percent of the cost up to $295, followed by 25 percent of the cost up $2,700, before entering the “donut hole” and paying all of the cost up to $6,154 (Joyce and Lau, 2009).

25 Consider as an example a diagnosis of high cholesterol. Someone who has slightly elevated cholesterol levels may benefit from a diet or lifestyle change over risking the negative side-effects of taking a statin whereas a person having very high cholesterol—perhaps with a history of heart disease—will have a higher prognosis with a statin than with a change in diet or lifestyle.
utility in health is concave, to insure that the firm’s problem is well-behaved we impose the technical restriction that \( \bar{u}(\theta) \) is a weakly convex function of \( \theta \).\(^{26}\)

Consumers are rational so the expected value of treatment contains an option value of all future expected consumption. Consumers and the firm have the common discount factor \( \delta \in [0, 1] \).\(^{27}\) When consumer \( i \) knows the efficacy of the drug then \( i \)’s total discounted expected utility of trying the drug is given as

\[
EU_i = (\phi \theta_i - p_1) + \phi \sum_{t=2}^{T} \delta^{t-1}(\theta_i - p_t) + (1 - \phi) \sum_{t=2}^{T} \delta^{t-1} \bar{u}(\theta_i).
\]
(1.1)

The first term of the consumer’s expected utility identifies the expected utility from the first period of treatment. The benefit is stochastic while the cost, \( p_1 \), is not. The second and third terms on the right-hand-side of (1.1) represent the option value of future consumption. A consumer will still purchase the drug for the first period of treatment when the first term is negative as long as the option value is sufficient to ensure the total discounted expected utility exceeds the consumer’s outside option.

When \( \phi \) is not commonly known, consumer \( i \)’s total discounted expected utility is given as

\[
EU_i = \int_{0}^{1} \left\{ (\phi \theta_i - p_1) + \phi \sum_{t=2}^{T} \delta^{t-1}(\theta_i - p_t) + (1 - \phi) \sum_{t=2}^{T} \delta^{t-1} \bar{u}(\theta_i) \right\} dG(\phi).
\]
(1.2)

Consequently, consumers’ expected utility from trying the drug has two dimensions of uncertainty. First, there is uncertainty regarding the overall efficacy of the drug.

\(^{26}\)Again, if we represent a consumer’s monetary utility for health with the concave function \( V \), and let \( f(\tau) \) be a linear transformation of \( \tau \) where \( 0 < f'(\tau) \leq 1 \) so that the outside option has some curative ability, but not more than the drug, then the outside option may be represented as \( \bar{u}(\theta) = V(H - f(\tau)) - V(H - \tau) \). The second derivative is then defined as \( \bar{u}''(\theta) = V''(H - f(\tau))|f'(\tau)|^2 - V''(H - \tau) \), which is positive for a large family of concave utility functions, \( V \), including those whose third derivative is zero or negative.

\(^{27}\)Assigning different discount factors to the consumers and firm will not change the results. The discount factor’s effect on the firm’s strategy is discussed in Section 1.3.
reflected by the distribution $G$, and second, there is uncertainty regarding the match outcome with any given individual reflected by the Bernoulli parameter $\phi$.

1.3 Prices

1.3.1 Prices without Sampling

Because the firm cannot individually price discriminate, the market is in a steady-state, and new consumers have a common prior, the spot price will remain constant from one period to the next ($p_t = p$ for all $t$), and the firm’s problem is simply that of choosing the spot price which maximizes its profit given the beliefs of the uninformed consumers and the distribution of illness severity.

Let $\mu$ represent the consumers’ common prior for the drug’s efficacy and define $\Sigma \equiv \sum_{t=0}^{T-1} \delta^t = (1 - \delta^T)/(1 - \delta)$. Plugging $\mu$ and $\Sigma$ into (1.2) and rearranging allows consumer $i$’s total discounted expected utility of trying the drug to be expressed as

$$EU_i = \mu \Sigma (\theta_i - p) + (1 - \mu) \left[ (\Sigma - 1) \bar{\pi}(\theta_i) - p \right].$$

(1.3)

The first term identifies the discounted present value of utility given the consumer is a match times the expected probability of being a match and the second term identifies the expected value of a mismatch.

An uninformed consumer, $i$, will purchase the drug if his expected utility from purchasing the drug is at least as high as his outside option, $EU_i \geq \Sigma \bar{\pi}(\theta_i)$. Using (1.3) the participation constraint can be expressed by consumer type as $\theta_i \geq \hat{\theta}_{NS}(p \mid \mu)$, where $\hat{\theta}_{NS}(p \mid \mu)$ solves

$$\hat{\theta}_{NS} = \left( \frac{1 + \mu(\Sigma - 1)}{\mu \Sigma} \right) \left[ p + \bar{\pi}(\hat{\theta}_{NS}) \right].$$

(1.4)
Thus, $\hat{\theta}_{NS}(p \mid \mu)$ identifies the type for which the expected value of trying the drug is equivalent to his outside option. Notably, the cutoff type is dependent on the belief, $\mu$, so any action the firm takes (or does not take) which affects the consumers’ belief will change the cutoff type.

The firm’s per period demand consists of the demand from new consumers together with the proportion of consumers who are a match with the drug and continue the treatment:

$$D(p) = (1 + \phi (T - 1)) [1 - F(\hat{\theta}_{NS}(p \mid \mu))]. \quad (1.5)$$

The firm’s discount factor does not affect the firm’s sampling decision given its infinite horizon and maximizing average profit is equivalent to maximizing the total discounted stream of profit:

$$(1 - \delta) \sum_{t=0}^{\infty} \delta^t \Pi_t = \Pi_t.$$  

Using (1.5), the firm’s average, expected profit given it does not provide samples is

$$E\Pi^{NS}(p; \phi) = \alpha p (1 + \phi (T - 1)) [1 - F(\hat{\theta}_{NS}(p \mid \mu))], \quad (1.6)$$

where $\alpha$ is the inverse of the consumers’ coinsurance rate.

The firm’s objective is to choose $p$ to maximize (1.6). The proof for Proposition 1 shows the firm’s optimization problem is concave; therefore, from the first-order condition, the optimal price $p^*_{NS}$ is the unique solution to the equation

$$p^*_{NS}(\mu) = \left( \frac{1 - F(\hat{\theta}_{NS}(p^*_{NS} \mid \mu))}{f(\hat{\theta}_{NS}(p^*_{NS} \mid \mu))} \right) / \hat{\theta}'_{NS}(p^*_{NS} \mid \mu). \quad (1.7)$$

The optimal price has the intuitive characteristics of increasing in the drug’s actual or expected efficacy, the expected number of periods of treatment and the discount
factor:

\[
\frac{\partial p^*_N}{\partial \mu} \geq 0, \quad \frac{\partial p^*_N}{\partial T} \geq 0, \quad \text{and} \quad \frac{\partial p^*_N}{\partial \delta} \geq 0.
\]

The comparative statics are a consequence of the option value of treatment. That is, the higher the option value of the treatment the higher the price the firm can charge.

1.3.2 Prices with Sampling

When the firm provides a sample, a consumer will initially try the drug if \( \theta_i \geq \bar{u}(\theta_i)/\mu \). Consumer \( i \) will also try the drug when \( \theta_i < \bar{u}(\theta_i)/\mu \); however, as long as the present discounted utility of doing so exceeds that of starting with the outside option (i.e., \( EU_i \geq \Sigma \bar{u}(\theta_i) \)). Once a consumer’s uncertainty has been resolved, the firm cannot gain by continuing to distribute samples.\(^{28}\) By removing the payment from the first period of consumption in Eq. (1.3), consumer \( i \)'s participation constraint can be expressed as

\[
\theta_i \geq \min \left\{ \bar{u}(\theta_i)/\mu, \left( \frac{\mu(\Sigma - 1)}{\mu \Sigma} \right) p + \left( \frac{1 + \mu(\Sigma - 1)}{\mu \Sigma} \right) \bar{u}(\theta_i) \right\}. \tag{1.8}
\]

Let \( \hat{\theta}_S \) denote the lowest type who tries a sample when available. We will refer to consumers as unconstrained whenever the RHS of (1.8) evaluates to \( \bar{u}(\hat{\theta}_S)/\mu \) at \( \hat{\theta}_S \). Consumers are unconstrained in the sense that regardless of their discount factor they need to compare the expected utility of trying the sample against the outside option.

\(^{28}\) This argument requires that the firm cannot observe the consumers’ type. Arguably a physician can observe a consumer’s type after the match value has been resolved and samples can be used to simulate first-degree price discrimination; however, if this occurred, we would observe higher rates of sampling for lower income and uninsured patients. Cutrona et al. (2008) perform the most comprehensive study using MEPS data and find that samples are far more likely to go to the insured. We do observe, however, in Tables 1.2 and 1.1 that some fraction of samples are given to existing patients suggesting that for some small group of patients, physicians use samples to lower their costs of treatment.
for the first period only, which is not the same as being myopic. Otherwise consumers must take into account their entire option value of trying the drug and we refer to them as *constrained* in this case.

After initially trying the sample, if the consumer is not a match then he will clearly not continue the treatment with the drug and switch to his outside option. If he is a match, then he will continue the treatment if and only if \( \theta_i - p \geq \bar{u}(\hat{\theta}_S) \).

To determine the firm’s revenue generating demand we start with the initial participation constraint. If consumers are *unconstrained*, then it must be the case that
\[
p > \left[ \frac{1 - \mu}{\mu} \right] \bar{u}(\hat{\theta}_S).
\]
In the second period of treatment, if
\[
p + \bar{u}(\theta_i) > \left( \frac{\mu(\Sigma - 1)}{\mu\Sigma} \right) p + \left( \frac{1 + \mu(\Sigma - 1)}{\mu\Sigma} \right) \bar{u}(\theta_i)
\]
for all \( \theta \geq \hat{\theta}_S \), then again it must be the case that \( p > \left[ \frac{1 - \mu}{\mu} \right] \bar{u}(\hat{\theta}_S) \). Therefore whenever consumers are *unconstrained* the continuation condition is \( \theta_i \geq p + \bar{u}(\theta_i) \). Otherwise, the initial participation and continuation conditions are equivalent. This follows because a consumer will only try the drug when \( \theta_i < \bar{u}(\theta_i)/\mu \) if his option value is sufficiently high to compensate for the expected loss in value in the first period and this occurs only when \( \theta_i > p + \bar{u}(\theta_i) \) for all \( \theta_i \geq \hat{\theta}_S \).

Let \( \hat{\theta}_S(p_S | \mu) \) denote the cutoff type when the firm provides samples, that is \( \hat{\theta}_S(p_S | \mu) \) solves
\[
\hat{\theta}_S = \begin{cases} 
p + \bar{u}(\hat{\theta}_S) \\
\left( \frac{\mu(\Sigma - 1)}{\mu\Sigma} \right) p + \left( \frac{1 + \mu(\Sigma - 1)}{\mu\Sigma} \right) \bar{u}(\hat{\theta}_S)
\end{cases}
\]
if consumers are *unconstrained*, and
\[
\hat{\theta}_S(p_S | \mu) \text{ otherwise.}
\]

\[(1.9)\]
Using the appropriate $\hat{\theta}_S(p | \mu)$ the firm’s total (revenue generating) per period demand can be expressed as

$$D(p_S) = \phi(T - 1)[1 - F(\hat{\theta}_S(p_S | \mu))]. \quad (1.10)$$

A firm cannibalizes the first period of demand when it provides new consumers free samples so all of the its revenue is from returning consumers. Proportion $\phi$ of the new consumers will have a successful match, and $1 - F(\hat{\theta}_S(p_S | \mu))$ of these will have a willingness to pay higher than the price, thus will purchase the drug for an expected total of $T - 1$ additional periods.

Using the per period demand in (1.10), the firm’s expected average profit when it provides free samples is

$$E\Pi^S(p_S; \phi) = \alpha p_S \phi(T - 1)[1 - F(\hat{\theta}_S(p_S | \mu))] - c_S, \quad (1.11)$$

where $\alpha$ is again the inverse of the consumers’ coinsurance rate.

By taking the first-order condition of (1.11) it can be shown that, because of the concavity of the firm’s profit function, $p^*_S > p^*_S$, regardless of whether or not consumers are constrained or unconstrained. The ability to optimally set a higher price naturally follows from the fact that consumers do not pay for the drug until after learning their match value, removing some of the risk of initializing treatment with the drug. When consumers are unconstrained, however, the firm will also optimally cover a larger proportion of the market despite setting a higher price.\(^{29}\) This follows because the cutoff type $\hat{\theta}_S$ does not require strictly positive surplus in every treatment.

\(^{29}\)When sampling also alters consumers’ beliefs the market will expand when consumers are constrained as well.
period starting with the second period. The following proposition summarizes these findings.

**Proposition 1.** By providing samples, a firm can optimally set a strictly higher price and cover more of the market for all $\phi \in (0, 1)$ and for all $\mu \in (0, 1)$; i.e.,

$$p^*_S > p^*_{NS},$$

$$\hat{\theta}_S^* \leq \hat{\theta}_{NS}^*.$$

*Proof.* See Appendix A

Note that the results of proposition 1 do not require a change in beliefs resulting from sampling; nor do they require myopia on the part of consumers.

In many cases the firm may not be able to adjust the price paid by consumers such as when consumers have insurance policies with strong prescription drug benefits. In this case sampling allows the firm to expand its market to lower types regardless of whether or not consumers are constrained. This follows as a direct corollary to Proposition 1.

**Corollary 1.** If the firm cannot adjust the price paid by consumers, then the firm will strictly cover more of the market for all $\phi \in (0, 1)$ and for all $\mu \in (0, 1)$ when it provides samples; i.e., $\hat{\theta}_S < \hat{\theta}_{NS}$.

*Proof.* See Appendix A

The market expanding ability of samples is what makes sampling an effective marketing tool and is reflected in studies such as Gönül et al. (2001) and Mizik and Jacobson (2004), which find that sampling increases sales for the sampled drug.
1.4 Equilibrium Sampling

In Sections 1.4.1 and 2.3.2 we analyze the firm’s optimal sampling and pricing decisions when the efficacy of its drug is commonly known, and known only to the firm, showing how the knowledge of a drug’s efficacy leads to different sampling strategies.

1.4.1 Symmetric Information

Naturally a firm will provide samples to a consumer if and only if the expected profit from providing samples is higher than the expected profit from not providing samples:

\[ E\Pi^S(p^*_S(\mu); \phi) > E\Pi^{NS}(p^*_{NS}(\mu); \phi). \]  

(1.12)

When information is symmetric, then \( \mu = \phi \); however, consumers still do not know ex ante if they are a match with the drug. By providing samples the firm allows consumers to learn their match before paying. Consequently the firm can charge the full-information price without holding up new purchases. Condition (1.12) can be expressed as

\[
p^*_S(\phi)(T - 1)[1 - F(\hat{\theta}_S(p^*_S(\phi) | \phi))] - c_S > 
\]

\[
p^*_{NS}(\phi)(1 + \phi(T - 1))[1 - F(\hat{\theta}_{NS}(p^*_{NS}(\phi) | \phi))].
\]

(1.13)

The left-hand-side of (1.13) is the profit the firm earns in one period if it provides samples and the right-hand-side is the profit the firm earns if it does not provide samples.\(^\text{30}\)

\(^\text{30}\)The factor \( \alpha \) is omitted from this expression and all further expressions. This is without loss of generality because \( \alpha \) simply serves to re-scale \( c_S \).
From inspection of (1.13) we can gain two immediate insights into the firm’s sampling decision under symmetric information. First, it is clear that, because there is a cost to providing samples, the following is true:

\[ E\Pi^S(p^*_S(0); 0) < E\Pi^{NS}(p^*_{NS}(0); 0). \]

Thus, because of the continuity of the profit functions, if there are efficacies for which the firm would like to optimally provide samples, then those must be higher efficacies. Second, it is also clear that when the drug systematically matches with everybody the firm can charge the full-information price \((p^*_S = p^*_{NS})\) and the following is also true:

\[ E\Pi^S(p^*_S(1); 1) < E\Pi^{NS}(p^*_{NS}(1); 1). \]

Thus, if there are efficacies for which the firm would like to optimally provide samples, then those efficacies must be in the interior of the interval \([0, 1]\). This makes intuitive sense, because, when the efficacy is common knowledge, a firm with a high efficacy drug can charge a price sufficiently close to the full information price and the potential gains to providing samples are small, hence eliminated by the cost of providing the samples. A drug with a high efficacy may be characterized as more of a search good than an experience good a lá Nelson (1974) because of the high degree of confidence consumers have in their match outcome. On the other hand, despite the ability to charge a higher price by providing samples, a firm with a drug having a lower efficacy will also choose not to provide samples. When a drug has a low efficacy, too few consumers will return to continue treatment, resulting in a loss in profit which exceeds the gain from the attendant higher price.
Increasing a drug’s efficacy has two effects on the firm’s no-sampling profit function. First, the firm can set a higher price with higher efficacy drugs because of the higher option value of treatment. Second, a higher efficacy results in a higher number of drug matches. The combination of these two effects cause the firm’s profit to be increasing and convex in efficacy. The efficacy has a similar effect on the sampling profit; however, the profit is less convex with efficacy and is linear with efficacy at the extreme when consumers are unconstrained. The convexity of the firm’s profit functions with respect to efficacy and the fact that at the end-points the firm prefers to not provide samples establishes the following proposition.

**Proposition 2.** Let $\Phi$ be the non-empty set of efficacies such that the firm provides samples for any $\phi \in \Phi$, then $\Phi$ is convex and $\Phi \subset [0, 1]$.

*Proof.* See Appendix A. \qed

It may be the case that it is never optimal to provide samples for any efficacy, i.e., $\Phi = \emptyset$. For example if the treatment length is sufficiently short, or the cost of sampling sufficiently high, then the firm will never find it optimal to provide samples as it will not be able to recover the cost of sampling from the higher price.

Interestingly, the higher the outside option is relative to the value of the drug, the larger the set of efficacies for which the firm will distribute samples. This follows because higher outside options increase the risk to the consumer of trying the drug, which increases the value of the sample. This suggests that sampling rates may be
higher for drugs having only marginally higher value from other treatment options.\textsuperscript{31}

This result is summarized in the following lemma.

**Lemma 1.** The sampling set of efficacies $\Phi$ is increasing with the outside option $\overline{u}(\cdot)$.

**Proof.** See Appendix A.

\hfill \square

### 1.4.2 Asymmetric Information

Before deriving the sampling equilibrium under asymmetric information, it is important to note that a firm cannot use the drug’s price to signal the drug’s efficacy. This limitation is important because, if a firm could signal the efficacy level of a drug through its choice of price, then the sampling equilibrium would be similar, if not identical, to the symmetric information equilibrium and the act of providing samples has no other effect on the equilibrium. We show, however, that the act of providing samples may function as a signal of higher efficacy resulting in a different sampling strategy.

**Proposition 3.** Under asymmetric information, a firm cannot signal the efficacy of the drug through price.

**Proof.** Suppose the firm can signal the value of $\phi$ with a price $p(\phi)$, then announcing a price is the same as announcing the efficacy, $\phi$. If the price announcement is to credibly signal the efficacy, then it must be the case that the firm cannot do better with an alternative price announcement. That is, if a firm’s drug has efficacy $\phi$, then

\textsuperscript{31}For example, this suggests that we should observe more sampling for “me-too” drugs that are not substantially better then existing drugs. To our knowledge there are no studies which identify the degree of sampling for me-too drugs relative to monopoly drugs.
an equilibrium must satisfy $E\Pi(p(\phi), \phi) \geq E\Pi(p(\hat{\phi}), \phi)$ for all $\phi, \hat{\phi} \in [0, 1]$ where $\phi \neq \hat{\phi}$, and

$$E\Pi(p(\phi); \phi) = p(\phi)(1 - \phi(T - 1))[1 - F(\hat{\theta}(p(\phi) \mid \phi))], \quad (1.14)$$

$$E\Pi(p(\hat{\phi}); \phi) = p(\hat{\phi})(1 - \phi(T - 1))[1 - F(\hat{\theta}(p(\hat{\phi}) \mid \hat{\phi}))]. \quad (1.15)$$

Therefore, using (1.14) and (1.15), $E\Pi(p(\phi); \phi) \geq E\Pi(p(\hat{\phi}); \phi)$ implies

$$p(\phi)[1 - F(\hat{\theta}(p(\phi) \mid \phi))] \geq p(\hat{\phi})[1 - F(\hat{\theta}(p(\hat{\phi}) \mid \hat{\phi}))], \forall \hat{\phi}, \phi \in [0, 1]. \quad (1.16)$$

However, if a firm’s drug has efficacy $\hat{\phi}$, then $E\Pi(p(\hat{\phi}); \hat{\phi}) \geq E\Pi(p(\phi); \hat{\phi})$ for all $\hat{\phi}, \phi \in [0, 1]$ implying

$$p(\hat{\phi})[1 - F(\hat{\theta}(p(\hat{\phi}) \mid \hat{\phi}))] \geq p(\phi)[1 - F(\hat{\theta}(p(\phi) \mid \phi))], \forall \hat{\phi}, \phi \in [0, 1]. \quad (1.17)$$

Both (1.16) and (1.17) are correct if and only if $p(\phi) = p(\hat{\phi})$ for all $\phi, \hat{\phi} \in \Phi$, a contradiction. Therefore price cannot credibly signal a drug’s efficacy.

The reason a firm cannot utilize its price to signal efficacy is because the firm does not incur a penalty for a misreport. In the introductory pricing literature the quality level affects the ex post valuation of the good; consequently, a firm setting a high price for a low quality good is penalized with a large loss in demand after consumers learn that the quality level is insufficient for the price. In the current model, the ex post match value of the drug is independent of the realization of $\phi$. Note that this result is not a consequence of the deterministic ex post match value. That is, if the ex post match value is modeled instead as a random variable, then the price can be used to credibly signal the mean of the distribution of match values conditional on
being a match. In this way, it is the *idiosyncrasy* in being a match which distinguishes pharmaceuticals from most other experience good markets.

When the firm is not able to provide samples, perhaps because of regulation, then the consumers’ beliefs will simply be the unconditional prior, \( \mu = \int_0^1 \phi \, dG(\phi) \). If, however, a firm has the ability to provide samples, then because the firm does not benefit from providing samples for all efficacies, its decision to dispense samples furnishes consumers with additional information with which to condition their belief. Because the firm’s optimal strategy must now take into account how its actions affect the consumers’ beliefs, and how those beliefs restrict what the firm can do, the appropriate equilibrium concept is a perfect Bayesian equilibrium.

**Definition 1.4.1.** A perfect Bayesian equilibrium of the sampling game consists of the tuple

\[
\langle s^*(\phi), \mu^*(s^*), p^*(\phi) \rangle,
\]

where \( s^* \in \{ S = \text{Sample}, NS = \text{No Sample} \} \) is the firm’s equilibrium sampling choice, \( \mu(s^*) \) is the consumers’ common equilibrium belief about the drug’s efficacy given \( s^* \), and \( p^* \) is the firm’s equilibrium price.

When the firm has the ability to provide samples, then the consumers’ prior is conditioned on the sampling decision of the firm. Define the consumers’ belief given the firm does not provide samples as \( \mu_{NS} = E[\phi \mid s = NS] \). Define \( \hat{\phi} \in [0,1] \) as the efficacy solving

\[
E\Pi^S(p^*_S(\mu_S); \hat{\phi}) = E\Pi^{NS}(p^*_S(\mu_{NS}); \hat{\phi}). \tag{1.18}
\]
That is, \( \hat{\phi} \) represents the efficacy level where the firm is just indifferent between providing samples and not providing samples. To see how the firm’s optimal strategy changes under asymmetric information, observe that the firm faces two problems that were not present when information is symmetric.

The first problem the firm encounters when information is asymmetric is that it cannot signal when the drug has a very high efficacy. Recall that when the efficacy is common knowledge, the firm finds it optimal to not provide samples for very high efficacy drugs. Because consumers cannot distinguish between when the drug has a very high efficacy from when the firm’s drug has a low efficacy when the firm does not provide sample, however, the firm must optimally provide samples for all high efficacy drugs. Consequently, when it exists, the cut-off efficacy \( \hat{\phi} \) will be unique, and the sampling strategy is monotonic in efficacy. The following lemma formally states this.

**Lemma 2.** The cut-off efficacy \( \hat{\phi} \) is unique, and the firm will provide samples for any efficacy above the cut-off.

The second problem the firm encounters when information is asymmetric is that the act of dispensing samples furnishes information about the efficacy. To see why this introduces a problem, define \( \lambda(\hat{\phi}) \) as the unique\(^{32} \) \( \lambda \in (0, 1) \) satisfying

\[
\lambda(\hat{\phi}) \hat{\phi} = E[\phi \mid \phi < \hat{\phi}] = \int_{0}^{\hat{\phi}} \phi \, dG(\phi) / G(\hat{\phi}).
\] (1.19)

The right-hand-side of (1.19) identifies the consumers’ expectation of the drug’s efficacy given the firm has not provided samples; i.e., given that the efficacy must be

\(^{32}\)Recall \( G \) has strictly positive density over the supports \([0, 1]\).
less than the cut-off $\hat{\phi}$. The coefficient $\lambda(\hat{\phi})$ can be interpreted as a measure of the degree to which not providing samples lowers the consumers’ belief; i.e., $\lambda$ is a measure of pessimism similar to Shapiro (1983). A heavier lower tail for the distribution $G$ results in a $\lambda(\hat{\phi})$ closer to zero. Consequently the price the firm can charge will naturally be lower when it is more likely that the drug has a low efficacy. Because the firm must now take the consumers’ beliefs into account when making its sampling decision, the firm finds it optimal to provide samples for lower efficacies than it would have if the consumers knew the efficacy. This argument is captured in the following lemma.

Lemma 3. The sampling cut-off efficacy is always lower under asymmetric information than under symmetric information; i.e., $\hat{\phi} < \bar{\phi}$, where $\bar{\phi}$ is the lower bound of the symmetric information sampling set.

However, as when information is symmetric, the firm may not find it optimal to provide samples for any efficacy in $[0,1]$. This again occurs when there is an excessive cost to sampling, there are too few number of periods of repeat purchase by the consumer, or the consumer does not discount the future too heavily permitting the firm to charge a high no-sampling price.

The following proposition combines lemmas 2 and 3 and formalizes the equilibrium.

Proposition 4. Let $\Phi$ denote the set of efficacies from which the firm will provide samples. If the efficacy, $\phi$, of a drug is not known by consumers, then there exists a unique perfect Bayesian equilibrium.
The equilibrium is partially separating when $\Phi \neq \emptyset$ and will have the following properties:

1. Monotone sampling: $s^* = S$ for all $\phi > \hat{\phi} > 0$ where $\hat{\phi}$ solves (1.18), otherwise $s^* = NS$.

2. Consumers have conditional beliefs

$$
\mu_{NS}^* = \frac{\int_{\hat{\phi}}^{\phi} \phi \, dG(\phi)}{G(\hat{\phi})} \quad \text{and} \quad \mu_S^* = \frac{\int_{\phi}^{1} \phi \, dG(\phi)}{1 - G(\phi)}.
$$

3. $p^*$ maximizes (1.11) when $s^* = S$ and maximizes (1.6) when $s^* = NS$.

The equilibrium is pooling when $\Phi = \emptyset$, $s^* = NS$, consumers have the unconditional belief $\mu_{NS} = \int_{0}^{1} \phi \, dG(\phi)$, and $p^*$ maximizes (1.6).

Proposition 4 identifies that the information asymmetry results in a monotonic sampling strategy in contrast to when information is symmetric. Though the ability of samples to alter beliefs likely has a second-order effect on prescription choice compared to being able to learn one’s match before purchasing the drug, Venkataraman and Stremersch (2007) provide evidence that efficacy is an important predictor of sampling by observing that samples are more likely to be provided for high efficacy drugs.

1.5 Sampling with Informed and Uninformed Consumers

An interesting extension to the analysis is to consider the firm’s sampling strategy when some consumers are better informed than others. For example, physicians may have different levels of experience or receive evaluative information about drugs from different sources such as studies published in medical journals or word of mouth.
from colleagues. Consequently some physicians may be better informed than others about any particular drug’s efficacy.

Pharmaceutical firms spend billions of dollars annually on physician detailing in which they discuss their firm’s drugs with physicians and it is reasonable to assume that the representatives develop a sense of the physicians’ beliefs regarding the efficacy of their drug through the detailing process. With this information the representative can decide whether or not to leave any particular physician with samples. Supporting the notion that pharmaceutical firms partially base their sampling decision on observable characteristics of physicians, Rabin (2007) observes that less experienced physicians are more likely to be given samples and Boltri et al. (2002) report that less experienced physicians tend to change their prescription choices more than experienced physicians when samples are present. If less experienced physicians are uncertain of a drug’s efficacy because of their lack of experience with a drug, then the firm may find it optimal to provide them samples. In contrast there may be no advantage to a firm providing experienced physicians samples when those physicians already know the efficacy of the drug.

To show how a mix of consumer knowledge alters the firm’s sampling decision, consider the following addition to the model. Suppose a proportion \( \gamma \) of the consumers know the efficacy of the drug—call them informed consumers—whereas the

---

33 Illustrating the volume of medical literature available to physicians to develop a prior for a drug’s efficacy, in a study on the role of information and drug diffusion, Chintagunta et al. (2009) collect 1,064 medical studies on Cox-2 inhibitors published from 1999-2005.
proportion $1 - \gamma$ do not know the drug’s efficacy but have the prior $G$—call them un-
informed consumers.\footnote{Without formally modifying the model to explicitly introduce physicians, we can think of the market as containing a unit mass of physicians where $\gamma$ are perfectly informed of the drug’s efficacy, and $1 - \gamma$ have a prior characterized by $G$. Every period a unit mass of new patients become sick and match with one physician so that there is a one-to-one mapping from physicians to new patients every period.} Capturing the assumption that pharmaceutical firms learn the physicians’ priors through the detailing process or through the physicians’ prescribing patterns, we assume that the firm knows each consumer’s type, but consumers do not know the other consumers’ prior, or their proportions. Let $\Phi_U$ denote the optimal sampling set for uninformed consumers, and let $\Phi_I$ denote the optimal sampling set for informed consumers. As lemmas 2 and 3 show, the sampling set of efficacies for uninformed consumers is larger than, and a superset of the sampling set for informed consumers; i.e., $\Phi_I \subset \Phi_U$.

Let $p_{NS}^j$ for $j \in \{U, I\}$ denote the optimal price the firm can charge either type of consumer given it does not provide samples. Instead of a simple sample/no-sample strategy space, because the sampling set for the less informed consumers is a super-
set of the sampling set for the informed consumers, the firm now has four feasible strategies which we define as follows:

1. \textit{Persistent Sampling} ($\phi \in \Phi^I$): The firm can choose to serve both types by pro-
viding samples to everybody and setting the price to $p^*_S$. The firm will utilize this strategy when it is optimal to provide samples under symmetric informa-
tion. A drug in this category will be heavily sampled throughout its product life. Examples may include drugs from the top four therapeutic categories, asthma and allergy remedies, anti-infective agents, analgesics and anti-inflammatory
medications, and antihypertensive drugs, which collectively account for more than 63 percent of all drug samples dispensed (Backer et al., 2000).

2. *Introductory Sampling* ($\phi \in \Phi^U$ and $\phi \notin \Phi^I$): The firm can choose to serve both types of consumer by providing samples to the uninformed consumers and setting the price to $p_{NS}^I$. The firm would select this option when it is not optimal to provide samples under symmetric information. Drugs in this category will exhibit heavy sampling when introduced to the market and the proportion of uninformed consumers is large, but the volume of sampling will decline over time as physicians gain experience with the drug and learn its efficacy. Drug sampling will also be higher with residents and physicians who are early in their career than more experienced physicians. The majority of drug sampling follows this strategy as for most therapeutic classes sampling is heaviest right after a new drug has been introduced to the market (Joseph and Mantrala, 2009).

3. *Established Pricing* ($\phi \notin \Phi^U$ and $(1-\gamma)E\Pi^{NS}(p_{NS}^I) > E\Pi^{NS}(p_{NS}^I)$): The firm can choose to not provide samples and only serve the consumers who know the efficacy by setting the price to $p_{NS}^I$. Because this strategy creates a hold-up problem for the uninformed, it will only be useful for well established drugs — e.g., older off-patent name brands — where the proportion of consumers who are uninformed is likely to remain low. The firm will select this strategy option when it is not optimal to provide samples under symmetric or asymmetric information. This corresponds to the *niche market* strategy of Bergemann and Välimäki (2006).
4. *Introductory Pricing* ($\phi \notin \Phi^U$ and $E\Pi^{NS}(p_{NS}^U) > (1 - \gamma)E\Pi^{NS}(p_{NS}^I)$): The firm can choose to not provide samples and serve both types by setting the price to $p_{NS}^U$. The firm’s choice would hinge on the size of the uninformed market and the differential between $p_{NS}^U$ and $p_{NS}^I$; i.e., the firm will choose this strategy over the previous strategy if $p_{NS}^U > \gamma p_{NS}^I$. As the proportion of informed consumers increases the firm’s strategy will flip over to the established pricing strategy. This corresponds to the *mass market* strategy of Bergemann and Välimäki (2006).

1.6 Welfare

Because the concern with sample dispensation has primarily been on how the practice affects consumers, we proceed by first focusing on consumer welfare before examining overall social surplus.\(^{35}\)

1.6.1 Consumer Surplus

When the sampling set, $\Phi$, is empty the firm does not provide samples for any realization of the efficacy. The fact that the firm does not provide samples provides no additional information to consumers and there is no change in consumer or producer surplus from permitting sampling. When the sampling set is nonempty, however, then consumer surplus is altered by permitting sampling. To see how, let $\mu_0$ represent the

\(^{35}\)The welfare analysis should be viewed as a baseline as there are important factors not explicitly modeled here that will also affect welfare. For example, there is some concern that sample package labeling and instruction is insufficient, thus resulting in unsafe drug consumption (Backer et al., 2000), reducing the consumer benefit to sampling. Moreover, because sample dispensation is more profitable for the firm, the practice increases the firm’s available funds for drug development. An increase in the firm’s profit will potentially increase the rate at which new drugs are developed and brought to market, increasing the consumer benefit to sampling.
consumers’ unconditional prior when sampling is not permitted; that is, let $\mu_0 \equiv \int_0^1 \phi \ dG(\phi)$. Suppose the efficacy of the firm’s drug is in the sampling set, then the per period change in consumer surplus from permitting sampling is

$$CS_S - CS_{NS} = \Delta U_1 + \phi(T - 1) \sum_{t=2}^{T} \Delta U_t$$

$$= \int_{\hat{\theta}_S}^{\theta_{NS}} \{ \phi \theta - \bar{u}(\theta) \} \ dF(\theta) + \int_{\hat{\theta}_{NS}}^{\bar{\theta}} p^*_S(\mu_0) \ dF(\theta)$$

$$+ \phi(T - 1) \left[ \int_{\hat{\theta}_S}^{\theta_{NS}} \{ \theta - p^*_S(\mu^*_S) - \bar{u}(\theta) \} \ dF - \int_{\hat{\theta}_{NS}}^{\bar{\theta}} \Delta p^* \ dF(\theta) \right],$$

where $\Delta p^* = p^*_S(\mu^*_S) - p^*_{NS}(\mu_0)$. The change in consumer surplus from permitting sampling consists of three components. The first component, represented by the first and third integrals of (1.20) identifies the consumer surplus derived from the market expanding effect of sampling. Absent samples these consumers would have started treatment with their outside option but the availability of samples have allowed these consumers to learn if they are a match first. The second component, represented by the second integral of (1.20), identifies the surplus change from the portion of the market covered when samples are not provided. With samples, these consumers no longer pay for their first period of consumption. Finally, the third component, represented by the last integral of (1.20), is the surplus loss by the portion of the market covered when samples are not provided and is a consequence of the higher price the firm can set when it dispenses samples. The first two components represent a gain in consumer surplus and the third represents a loss.

Eq. (1.20) can be rearranged to derive the following result.
Proposition 5. Suppose a monopolist’s drug has unknown efficacy $\phi$ and $\Phi$ is the sampling set. Consumer welfare is increased by permitting sampling under the following conditions:

1. When $\phi \in \Phi$, consumer welfare is increased if and only if
   \[
   \phi T \int_{\hat{\theta}_S}^{\hat{\theta}_S^*} \{\theta - \pi(\theta)\} \ dF(\theta) > \\
   p_S^*(\mu_S^*)(T - 1)[1 - F(\hat{\theta}_S^*)] - p_{NS}^*(\mu_0)(1 + \phi(T - 1))[1 - F(\hat{\theta}_{NS}^*])
   \] (1.21)

2. When $\phi \notin \Phi$, welfare is unchanged when $\Phi$ is empty and welfare is always increased when $\Phi$ is nonempty.

The left-hand-side of (1.21) represents the consumer surplus derived from the market-expanding effect of sampling and the right-hand-side is the difference in the firm’s revenues from providing samples and the firm’s revenue when sampling is not permitted. In words, proposition 5 says that when the firm chooses to provide samples, consumer surplus is increased if and only if the consumer surplus derived from market expansion exceeds the change in the firm’s revenue from allowing it to provide samples. When consumers are constrained sampling does not have a market expanding effect and consumer surplus is lowered with sampling. However, if the firm is unable to adjust the price consumers pay by dispensing samples, then per corollary 1 sampling expands the market as the firm is able to cover lower types, which may increase consumer welfare considerably. Given that some consumers have generous prescription drug benefits and others do not, sampling can have the perverse effect of decreasing the welfare of those who pay out-of-pocket while increasing the welfare for those who are covered by insurance.
The uninsured and indigent are one group who have substantial out-of-pocket costs. However, some physicians may provide sufficient samples to indigent patients that they can complete their entire treatment and this represents behavior not directly captured by the model. The overall welfare effect of sampling will also depend on the number of indigent patients receiving free treatment via samples. To our knowledge, nobody has yet looked at the number of patients who have received treatment via free samples as the literature has primarily focused on what proportion of those receiving free samples are poor, or without insurance (See, for example, Cutrona et al. (2008)).

We have so far discussed the factors contributing to increasing the left-hand-side of (1.21); however, note that the firm may find it optimal to provide samples when permitted to do so, but receive a lower revenue than when sampling is not permitted, making the right-hand-side of (1.21) negative. This counterintuitive result may occur because, when sampling is permitted but the firm does not provide samples, consumers infer the drug has a low efficacy and the firm is restricted to setting the price \( p_{NS}^*(\mu_{NS}) \), which is less than the optimal no-sampling price based on the unconditional prior, \( p_{NS}^*(\mu_0) \). Sampling may thus raise the firm’s profit over the profit it can earn with \( p_{NS}^*(\mu_{NS}) \) but not over the profit it earns with \( p_{NS}^*(\mu_0) \). When the firm’s revenue decreases, consumer surplus is increased by sampling regardless of whether or not sampling expands the market. If information is symmetric, then the no-sampling revenue can never be less than the sampling revenue when the firm optimally chooses

\[36\] This is one reason many physicians favor receiving free samples (Peterson et al., 2004).

\[37\] Recall that the cost of sampling, \( c_S \), implicitly accounts for any excess samples given to indigent patients to cover their entire course of treatment; a quantity the firm is assumed to know.
to provide samples. In this case consumer welfare is increasing only if the surplus generated by expanding the market exceeds the firm’s increase in revenue.

1.6.2 Socially Efficient Sampling

If a social planner is responsible for determining whether or not a firm should provide samples, then what would the planner choose? If the social planner is a hospital administrator interested in maximizing consumer welfare, then the planner will permit sampling only when the conditions of Proposition 5 are met. If the planner’s interest is in maximizing total social surplus, however, then she may still choose to permit sampling, even if it lowers consumer surplus.

The change in producer surplus from permitting sampling is given as follows:

\[
\Delta \Pi = \begin{cases} 
E \Pi^S(p^*_S(\mu^*_S); \phi) - E \Pi^{NS}(p^*_S(\mu_0); \phi) & \text{if } \phi \in \Phi, \\
E \Pi^{NS}(p^*_S(\mu^*_NS); \phi) - E \Pi^{NS}(p^*_S(\mu_0); \phi) & \text{if } \phi \notin \Phi. 
\end{cases}
\]

When \( \phi \in \Phi \), the change in producer surplus is the difference between the profits from providing samples and not providing samples. When the firm does not provide samples, the change in producer surplus comes from the difference in the price the firm can charge. If the conditional prior, given that the firm does not provide samples \( \mu^*_NS \) is less than the unconditional prior \( \mu_0 \), then producer surplus will be lower. Using the definitions of \( E \Pi^S \), \( E \Pi^{NS} \), and the change in consumer surplus identified in (1.20) it is straightforward to derive the following result.

**Proposition 6.** Suppose a monopolist’s drug has unknown efficacy \( \phi \) and \( \Phi \) is the sampling set. If \( \phi \in \Phi \), then social surplus is increased by sampling if and only if

\[
\phi T \int_{\hat{\theta}_S}^{\hat{\theta}^*_NS} \{ \theta - \bar{\pi}(\theta) \} \, dF(\theta) > c_S.
\]
Otherwise permitting sampling is socially wasteful. If \( \phi \notin \Phi \), then there is no change in social surplus.

When a drug’s efficacy is in the sampling set \( \Phi \), then any increase in social welfare is a result of welfare gained by consumers in the expanded market. It is thus important to identify the major strategic effects of sampling. If sampling primarily allows the firm to adjust the price consumers pay, for example when consumers are constrained, then there will be no market expansion and sampling is socially wasteful. If, however, the firm is unable to adjust the price consumers pay because of generous insurance benefits, then again, per corollary 1, sampling has the potential to greatly expand the market, which, if sufficiently large, will be socially beneficial.\(^{38}\)

When the efficacy of the drug is not in the sampling set, there is no change in social welfare. Consumers infer a lower level of efficacy for the drug given the firm does not provide samples resulting in a lower price. The gain in consumer surplus from the lower price is a zero-sum transfer from the firm to consumers.

1.7 Discussion and Conclusion

We have developed a simple, stylized model of sampling as a dynamic game of incomplete information. By assuming samples are used to mitigate the consumers’ uncertainty in their match outcome and by solving for the firm’s sampling strategy under symmetric and asymmetric information regimes the model generates qualitative results consistent with observations found in the medical literature. For example the model provides insight into why sampling rates vary substantially across therapeutic

\(^{38}\)The difference in the benefit of the drug and the full price paid by the insurer may exceed that of the outside option, in which case sampling will be socially wasteful.
classes, why sampling rates can vary substantially over a drug’s lifecycle with most taking place immediately after market introduction, why sampling is targeted more at early career physicians, and how sampling can alter a physician’s prescription choice. Moreover, the model predicts higher rates of sampling for lower differences in expected drug benefit over an outside option suggesting that we should observe higher rates of sampling for “me-too” drugs having similar value to first-to-market drugs.

The model shows that the idiosyncratic nature of the consumer-to-drug match is the distinguishing characteristic that separates prescription drugs from other experience goods. Given the low to moderate efficacy rate of many drugs, sampling provides a way for consumers to learn their idiosyncratic match with a drug before paying for treatment. When the efficacy of the drug is not commonly known, as would be the case early in a drug’s life-cycle, the firm dispenses samples to facilitate learning, which has a secondary consequence of signaling that a drug exceeds some cut-off efficacy. The signaling aspect of sampling under asymmetric information alters the firm’s strategy in two ways. First, the firm finds it necessary to provide samples for all high efficacy drugs, even though it would not do so when information is symmetric. This results in a monotonic sampling strategy where above some cut-off efficacy the firm dispenses samples. And second, the signaling aspect of sampling causes the firm to optimally provide samples for drugs having a lower efficacy than it otherwise would when information is symmetric. Consequently the sampling set of efficacies is larger under asymmetric information. Using this property of the sampling sets, the

39Sampling is prevalent in other product markets as well such as textbooks; however, in these markets sampling is better thought of as a form of quality revealing advertising as issues stemming from not being a match and foregoing alternative treatments are nonexistent.
firm’s pricing and sampling strategies were identified when the market exhibits a mix of knowledge.

Additional empirical research should be performed to identify if there are other important characteristics of therapeutic classes or physicians that drive sample usage not captured in this model. For example, are a physician’s beliefs and characteristics more important than a patient’s? Do some physicians exhibit a higher willingness to experiment with more treatments, or are less likely to prescribe without first being able to use samples because of risk aversion, and to what degree do patients concur with their physicians and their physicians’ beliefs, particularly when they have different risk preferences? What characteristics of these physicians caused such dispensation differences? We leave these questions to future research.

\[40\] For example, Backer et al. (2000) find significant differences in physician sample dispensation behavior. In one particular two-physician practice they found that one physician dispensed samples in 41.9 percent of all observed patient encounters whereas the other physician only dispensed samples in 3.2 percent of the observed patient encounters.
<table>
<thead>
<tr>
<th>Therapeutic Category</th>
<th># R</th>
<th>Prescription (R) Sample Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(Millions)</td>
<td>All R</td>
</tr>
<tr>
<td>Cardiovascular</td>
<td>746</td>
<td>0.0132 (0.0004)</td>
</tr>
<tr>
<td>Central Nervous Sys.</td>
<td>523</td>
<td>0.0148 (0.0005)</td>
</tr>
<tr>
<td>Metabolic</td>
<td>517</td>
<td>0.0244 (0.0007)</td>
</tr>
<tr>
<td>Hormones</td>
<td>248</td>
<td>0.0226 (0.0010)</td>
</tr>
<tr>
<td>Psychotherapeutic</td>
<td>206</td>
<td>0.0297 (0.0013)</td>
</tr>
<tr>
<td>Respiratory</td>
<td>188</td>
<td>0.0312 (0.0013)</td>
</tr>
<tr>
<td>Gastrointestinal</td>
<td>158</td>
<td>0.0311 (0.0014)</td>
</tr>
<tr>
<td>Topical Agents</td>
<td>125</td>
<td>0.0470 (0.0020)</td>
</tr>
<tr>
<td>Nutritional Products</td>
<td>96</td>
<td>0.0202 (0.0014)</td>
</tr>
<tr>
<td>Coagulation Modifiers</td>
<td>56</td>
<td>0.0157 (0.0018)</td>
</tr>
</tbody>
</table>

Table 1.1: 10 Most Prescribed Therapeutic Categories, 2008

Notes: Prescription data come from the 2008 Medical Panel Survey. Frequencies report the proportion of prescriptions with which a free sample was provided and does not indicate the number of samples provided per prescription. Number of prescriptions indicates the weighted number of prescriptions in millions. Standard deviations are reported in parenthesis. Frequencies and standard deviations are unweighted.
<table>
<thead>
<tr>
<th>Drug Name</th>
<th># Drugs in Class</th>
<th>Prescription (R) Sample Frequency</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>All R</td>
<td>New R</td>
<td>Existing R</td>
<td></td>
</tr>
<tr>
<td>Celebrex</td>
<td>3</td>
<td>0.2115</td>
<td>0.3393</td>
<td>0.1105</td>
<td></td>
</tr>
<tr>
<td>Yaz</td>
<td>144</td>
<td>0.1939</td>
<td>0.1852</td>
<td>0.1563</td>
<td></td>
</tr>
<tr>
<td>Nasonex</td>
<td>25</td>
<td>0.1709</td>
<td>0.2590</td>
<td>0.0802</td>
<td></td>
</tr>
<tr>
<td>Nasacort AQ</td>
<td>25</td>
<td>0.1605</td>
<td>0.3000</td>
<td>0.0227</td>
<td></td>
</tr>
<tr>
<td>Cymbalta</td>
<td>120</td>
<td>0.1590</td>
<td>0.2208</td>
<td>0.1111</td>
<td></td>
</tr>
<tr>
<td>Lexapro</td>
<td>51</td>
<td>0.1497</td>
<td>0.3222</td>
<td>0.0865</td>
<td></td>
</tr>
<tr>
<td>Crestor</td>
<td>37</td>
<td>0.1475</td>
<td>0.2340</td>
<td>0.1081</td>
<td></td>
</tr>
<tr>
<td>Lyrica</td>
<td>10</td>
<td>0.1404</td>
<td>0.2083</td>
<td>0.0579</td>
<td></td>
</tr>
<tr>
<td>Advair Diskus 250/50</td>
<td>29</td>
<td>0.1295</td>
<td>0.2083</td>
<td>0.1022</td>
<td></td>
</tr>
<tr>
<td>Vytorin</td>
<td>7</td>
<td>0.1262</td>
<td>0.2444</td>
<td>0.1055</td>
<td></td>
</tr>
</tbody>
</table>

Table 1.2: 10 Most Sampled Prescription Drugs, 2008

Notes: Prescription data come from the 2008 Medical Panel Survey. Only drugs which have been prescribed to at least 35 patients are included in the list. Frequencies report the proportion of prescriptions with which a free sample was provided and does not indicate the number of samples provided per prescription. Standard deviations are reported in parenthesis. Frequencies and standard deviations are unweighted. Patent indicates whether or not the drug is still on-patent.
<table>
<thead>
<tr>
<th>Drug Name</th>
<th>Patent Left</th>
<th># Years of Prescription (R) Sample Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>All R</td>
</tr>
<tr>
<td>Crestor</td>
<td>8</td>
<td>0.1475</td>
</tr>
<tr>
<td>Vytorin</td>
<td>6</td>
<td>0.1262</td>
</tr>
<tr>
<td>Lipitor</td>
<td>3</td>
<td>0.0784</td>
</tr>
<tr>
<td>Tricor</td>
<td>3</td>
<td>0.0744</td>
</tr>
<tr>
<td>Lovastatin</td>
<td>-</td>
<td>0.0219</td>
</tr>
<tr>
<td>Simvastatin</td>
<td>-</td>
<td>0.0181</td>
</tr>
</tbody>
</table>

Table 1.3: Most Sampled Statins, 2008

Notes: Prescription data come from the 2008 Medical Panel Survey. Only drugs which have been prescribed to at least 35 patients are included in the list. Frequencies report the proportion of prescriptions with which a free sample was provided and does not indicate the number of samples provided per prescription. Standard deviations are reported in parenthesis. Frequencies and standard deviations are unweighted. # years of patent left are from 2008. Both Lovastatin and Simvastatin were off-patent generics in 2008.
Chapter 2: Regulating a Monopolist with Unverifiable Quality

In early 2010 the U.S. Congress passed the Patient Protection and Affordable Care Act (PPACA) in an effort to increase access, lower costs, and improve quality for health care. One component of the bill provides funds for the Centers for Medicare and Medicaid Services (CMS) to research, develop, and test new payment and delivery arrangements for health care providers to achieve the latter two objectives. Interestingly, the PPACA already approaches the regulation of health insurers in a similar manner to the regulation of public utilities. For example, the PPACA contains strong language regarding the ratemaking and degree of coverage for health insurers participating in health exchanges.

Treating health insurers or providers as public utilities is appealing due to the large body of knowledge and experience in their regulation; however, the regulatory approaches taken for public utilities may not be completely transferable to health markets. For example, in telecommunications quality can be partially identified by quantitative measures such as the time to connect or the drop call ratio, and with public water quality can be identified by the quantitative measure of contaminant parts per million. In contrast, in health markets quality may refer to treatment
techniques, intensities, or technological sophistication that cannot be easily defined or measured, even if observable by a regulator. Consequently, minimum service quality regulation may be impractical or undesirable for health services. Moreover, rate-of-return regulation may also be impractical in such markets where the cost of providing the good or service to an individual is easily disguised due to consumer heterogeneity and economies of scope.

Similar limitations to regulating quality also arise in other markets such as public and higher education. For instance, the 2008 Charter School Renewal Quality Review Handbook for the Oakland Unified School District provides an itemized list of characteristics it uses to measure the quality of schools. The difficulty in quantifying quality levels, however, is reflected in the following statement from the handbook report (emphasis added):

It is also imperative that everyone recognizes that there are many ways in which a school’s program for improving student outcomes can merit a particular evaluation and that awarding levels is a matter of informed professional judgment and not simply a technical process.

Given the unique challenges quality introduces to a regulatory environment, the purpose of this study is to take a new look at the design of regulatory policy for a firm that can manipulate demand through its choice of unverifiable quality. Traditionally the mechanism design literature on regulating firms with unverifiable quality assumes demand is inelastic to quality, or has modeled quality as if it is a second good or choice variable. For example, in their treatment of quality Laffont and Tirole (1986) consider models where quality replaces effort and models having both quality and effort but where quality can be substituted with a linear transformation of quantity, which because it is not a choice variable for the firm, eliminates the problem of unverifiability.
then the firm really has only one choice variable: the quantity of the good that it wants to sell. The quality level of the good is simply a function of this choice. Reflecting the inseparability of quality and quantity, we transform the firm and regulator’s problems to make this relationship explicit; that is, despite the presence of quality, the firm has only one choice variable: the quantity of output. This approach more realistically captures the relationship between quantity and quality in many markets and produces substantially different outcomes than those found in the previous literature.

Markets in which consumer demand is affected by quality—such as health services, insurance, and education—generally contain a mix of for-profit and not-for-profit firms. Therefore, the firm’s objective function in this chapter is represented as a weighted sum of profit and community-benefit. This enables a comparison of how the regulator’s problem is affected by the firm’s objectives. The objective function admits as special cases a pure profit-maximizer and an altruistic output-maximizer for easy comparison with the previous literature. Additionally, the consumers’ access to the good is altered by considering a scenario where consumers are responsible for paying for their consumption directly and a scenario where the regulator pays on behalf of consumers using funds raised via taxation. Adopting the terminology of Caillaud et al. (1988), we refer to the good in the former scenario as a marketed good, and in

\[\text{For example, of the 4,897 registered community hospitals in the United States, 873 are for-profit while 2,913 are non-profit (AHA fast facts on US hospitals: http://www.aha.org/aha/resource-center/Statistics-and-Studies/fast-facts.html), and twelve states permit for-profit corporations to operate charter schools (National Education Association: http://www.corpwatch.org/article.php?id=886).}\]
the latter as a *nonmarketed good*. The marketed good represents the classical regulatory environment generally associated with public utilities. The nonmarketed good represents the regulatory environment most often attributed to markets for health care, but is relevant for any market in which the government is responsible for the provision of the good. Examples in the U.S. where the government is responsible for the payment, but not the production of the good, include public education provided by charter schools, voter registration services, military contracting, and of course, healthcare through Medicare, Medicaid, and the State Children’s Health Insurance Program (SCHIP).

As has long been understood in Bayesian mechanism design problems, the firm is often able to extract an information rent in the presence of asymmetric information; however, the contribution to the distortion away from the social optimum caused by the interaction of the firm’s and consumers’ best-response to the regulator’s payment rules has not been thoroughly explored. For example, the key finding of Lewis and Sappington (1988a) is that, when there is asymmetric information about the demand state, the firm is unable to benefit from its informational advantage and the regulator can induce the socially optimal output. This result is in sharp contrast to when the asymmetric information is with respect to the firm’s cost. In this case, Baron and

---

43Seminal works include Myerson (1979), Baron and Besanko (1984), Laffont and Tirole (1986), Sappington (1982), Riordan (1984) and Lewis and Sappington (1988a)


45Armstrong and Sappington (2004) have taken a first step in providing a synthesis of the regulatory problem. Their focus, however, is on adverse selection and how it affects the regulator’s payment rules and the market outcomes and do not consider the additional problems created by moral hazard.
Myerson (1982) find that the firm is able to extract an information rent resulting in an output that is always distorted down from the socially optimal level.

The contrasting findings of these two studies highlight the fact that asymmetric information is not a sufficient condition for a distortion away from the social optimum. Moreover, moral hazard does not necessarily lead to distortions either. Caillaud et al. (1988) first note this with an example of hidden effort and unobserved firm cost. There is no distortion away from the social optimum in this case because the agent internalizes all of the gains from exerting effort. The agent consequently exerts effort up to the point that the marginal benefit of additional effort is equal to the marginal cost, which is exactly the social optimum. When the agent is reimbursed based on an observable cost, then its cost savings from exerting effort are not internalized and the agent’s incentives are altered in such a way that the agent’s choice of effort is distorted away from the social optimum. Examples of such distortions abound in Laffont and Tirole (1993) who utilize a framework of compensation based on observable costs.46

Unlike firm effort, unverifiable quality represents a dimension of moral hazard that affects consumers. Consequently, quality can result in a distortion away from the social optimum without any adverse selection (Baron, 1981; Spence, 1975). This distortion occurs because of a conflict in the response by the firm and consumers to the regulator’s policy instrument and the fact that the firm can manipulate consumer demand through its choice of quality. This difference highlights the importance of

46See also Laffont and Martimort (2002) and Armstrong and Sappington (2004) for references to related models. In a similar model to the one here, when the market is characterized as having a nonmarketed good and symmetric information, Ma (1994) also observes that prospective payment provides a hospital with the appropriate incentive to choose the socially efficient level of effort.
accounting for both the consumers’ and firm’s response to the contract design. Another interesting characteristic of utilizing quality as the source of moral hazard is that the firm’s optimal choice of quality is directly linked to the state parameter (i.e., the source of adverse selection), thus allowing us to study how the linkage between the two affect the contract.

We generate two principal findings. First, in a regulatory environment containing a marketed good, the results of the model show that one cannot predict \textit{a priori} what form the output distortion will take when the firm’s choice of non-contractable quality influences demand, even for a pure profit-maximizing firm. Depending on the relative price- and quality-elasticities of demand and cost, a firm’s rents may be either increasing or decreasing with the unit payment resulting in a distortion of output that results in either an over- or under-supply relative to the socially-optimal level. This finding is in marked contrast to the familiar under-supply relative to first-best outcome obtained in similar models that either do not include quality (implicitly setting the quality-elasticity of demand to zero), explicitly set quality elasticity to zero, or have observable firm costs. When the good is nonmarketed, however, the distortion away from the social optimum is similar to the earlier literature because the price-elasticity of demand is forced to zero and output is uniquely determined by the quality level.

Our second finding is that the firm’s informational advantage does not necessarily result in an output distortion. Not surprisingly, we find that the second-best payment policy takes on a very different form depending on the objective of the firm and the consumers’ access to the good. Somewhat surprisingly, however, we find that
the regulator can completely attenuate the firm’s informational advantage when the
good is nonmarketed and the firm’s preference for community benefit over profit is
sufficiently strong that the firm acts as a pure output-maximizer. Note that the firm
is not a perfect agent of the regulator as in Ellis and McGuire (1986), but rather the
nature of the firm’s objective allows the regulator to more precisely control the firm’s
choice of output with the available policy instruments. This finding suggests that
in those regulated markets that include a mix of for-profit and not-for-profit firms,
the regulator’s payment policies must be tailored carefully with the firm’s objectives
if it is to achieve an optimal outcome; furthermore, the optimal outcome may differ
substantially depending on the firms’ objectives. Offering the two types of firm the
same contract, as is currently done with Medicare reimbursements for example, is
clearly sub-optimal. To eliminate the firm’s informational advantage, the regulator
faces a trade-off in reducing the firm’s informational advantage and incurring the
social deadweight loss from raising public funds. Thus, the analysis can also be
thought of as providing insight into when it is beneficial to utilize public funds to
provide access to a good or service and when it is not; e.g., a single-payer versus
market oriented healthcare payment system.

We utilize the standard techniques of Bayesian mechanism design so the mechan-
ics of the chapter are the same or similar to those found in many papers studying the
regulation of firms (e.g., Baron and Myerson (1982), Laffont and Tirole (1986), Cail-
land et al. (1988), and Lewis and Sappington (1992)). The information framework of

47In a recent study Landon et al. (2006) examined the quality of care for myocardial infarction,
congestive heart failure, and pneumonia provided by for-profit and nonprofit hospitals and, consistent
with the findings of this chapter, find that patients were more likely to receive higher quality care in
a nonprofit hospital reflecting the differences in outcome that result from the same payment scheme.
our model is closest to Lewis and Sappington (1992) who study the design of incentive programs to induce public utilities to provide a basic service with enhancements. Lewis and Sappington (1992) partially analyze the optimal contracts when quality enhancement is observable, the quantity consumed is observable, and when neither are observable. They do not consider other incentive regimes such as nonmarketed goods or mixed-objectives firms, nor do they fully characterize the distortions from the social optimum.

This model is also similar to Baron and Myerson (1982) and Lewis and Sappington (1988a) with respect to the regulator’s information. In Baron and Myerson (1982) the firm has superior knowledge of its costs, and in Lewis and Sappington (1988a) the firm has superior knowledge of the demand; however, neither model considers moral hazard, quality or otherwise and we show that neither is robust to the inclusion of quality. Lewis and Sappington (1988b) expand on Lewis and Sappington (1988a) by adding a second dimension of adverse selection: asymmetric information in the firm’s cost in addition to the market demand. Although only a single dimension of adverse selection is considered here, because quality affects demand, asymmetry in cost creates asymmetry in demand information and, likewise, asymmetry in demand creates asymmetry in cost. This relationship between cost and demand generates consistent outcomes between the two sources of asymmetric information. The results of Lewis and Sappington (1988b) would be similar to some of the results here if, in their model, the adverse selection parameters were perfectly correlated with one another.
Aguirre and Beitia (2004) modify the model of Lewis and Sappington (1988a) by making the consumer responsible for the unit payment, and the regulator responsible for the transfer payment. We also modify the source of payment to explore the role of the consumer incentive response to the regulator’s payment policy by considering a marketed and nonmarketed good, but either the consumer or the regulator is responsible for both payments in each case. When the payment is split, the regulator has a strict preference for unit payments over the fixed transfer because of the deadweight loss attributed to raising public funds. This preference for one payment over the other prevents the regulator from achieving the socially optimal outcome and, as such, is a step backwards for the regulator, who can achieve the socially optimal outcome if either party is responsible for both payments.

Finally, this chapter also relates to the literature that looks at optimal cost sharing rules for a firm having a partially altruistic objective. Ellis and McGuire (1986) first consider a firm with similar preferences as the regulator. The authors show that when the preferences of the firm are the same as the regulator’s, then the appropriate payment rule will induce the firm to produce the socially preferred level of quality. If the firm has less of a preference for consumer benefits, however, then the regulator must subsidize the firm in order to induce the socially preferred level of quality. Chalkley and Malcomson (1998a) introduce the possibility of cost-reducing effort by the firm. They find that as long the firm gains some value from supplying quality then an appropriate fixed payment is sufficient to induce the socially optimal level of effort, similar to Caillaud et al. (1988), but at a sub-optimal level of quality. Shifting to cost reimbursements will improve quality, increasing welfare. Finally, Jack (2005)
introduces uncertainty in the degree of altruism for a firm and derives the optimal cost-sharing rules when quality is observable, unobservable, and when the firm has an unknown degree of altruism.

There are three ways in which these analyses differ from this chapter. First, costs are assumed to be observable so the regulator utilizes a cost sharing rule to induce the firm to produce, which results in sub-optimal cost-reducing effort when effort is included in the model. Second, demand is inelastic to quality, which by itself simplifies the regulator’s problem, but also requires that any incentive to produce quality is completely motivated by altruism. Lastly, in this chapter’s model the firm is unable to earn negative profits which substantially alters the analysis when the firm has a sufficiently strong preference for community benefit.

The derivation of the our findings progresses as follows. In Section 2.1 the primary model is developed, including the derivation of the quality-adjusted cost and demand functions. In section 2.2 we establish the base-line, socially-optimal contract. In Section 2.3 we analyze the regulator’s problem when it chooses to have consumers pay directly for the good or service, and in Section 2.4 we analyze the regulator’s problem when it pays for the good on behalf of consumers. We establish the robustness of the results to asymmetric demand information in Section 2.5. Finally, in Section 2.6 we summarize the findings of the chapter and discuss some applications.
2.1 The General Model

Consumers

Consider a market environment where there is a single firm supplying a good or service at some level of quality \( q \in \mathbb{R}_+ \). Quality is observable by consumers but is not verifiable, so it cannot be directly contracted upon.\(^{48}\) We start by assuming that the firm has superior knowledge of its cost and later will show that the results are robust to asymmetric information about consumer demand. Demand for the firm's good, \( x(p,q) \), is a function of the price \( p \) and the quality, \( q \), of the good.\(^{49,50}\) Consumer demand can also represent a residual demand function for an imperfectly competitive market as with the market for hospital services, which includes substantial spatial competition. Under this interpretation, \( x(p,q) \) represents the firm's residual demand given all other firms supply some equilibrium level of quality and the regulator is assumed to take the number of firms in the market as given.\(^{51}\) Demand is \( C^2 \), increasing

\(^{48}\)The quality attribute may capture different characteristics depending on the market. For example in markets for health services quality may be some measure of the length of stay, number of hospital-induced complications, staff per patient, and in education it may reflect the expertise of the teachers or college admission rates. In any case, quality represents a characteristic that cannot be varied on a consumer-by-consumer basis.

\(^{49}\)It is not necessary that consumers perfectly observe quality as long as the consumers’ response to a change in demand is differentiable and predictable by the firm. Moreover, in a setting such as for hospital services, the relevant observer of quality may be a health maintenance organization (HMO), which upon observing the quality of a hospital makes the decision of whether or not to add the hospital to its network, thus affecting the demand for the hospital's services. Supporting the notion that some attributes of quality are observed by HMOs, Gowrisankaran and Town (2003) present evidence that competition between hospitals increases the quality of care for HMO patients.

\(^{50}\)Chalkley and Malcomson (1998a) consider the optimal regulatory policy when consumers cannot observe socially valuable quality but the hospital has an altruistic motive, without which the firm would always supply the lowest level of quality.

\(^{51}\)See Wolinsky (1997), Auriol (1998), Gravelle (1999), and Beitia (2003) for examples of models which specifically utilize structured competition to regulate unverifiable quality.
and concave in $q$, and decreasing in $p$. The regulator cannot observe the quantity of the good or services consumed, otherwise, knowing the price, the regulator will know the quality level.

Gross consumer benefit is represented by the function $B(x, q)$. $B$ may reflect the consumers’ direct value of consumption (i.e., $B(x, q) = \int_0^x P(\tilde{x}, q) d\tilde{x}$ where $P$ is inverse demand) or $B$ may reflect the regulator’s valuation of consumption in the presence of social externalities as are common with services such as health care and education. Gross consumer benefit is increasing and concave in the quantity and quality; $B_x > 0$, $B_q > 0$, $B_{xx} \leq 0$, and $B_{qq} \leq 0$. Moreover, marginal consumer benefit is weakly increasing in quality, $B_{xq} \geq 0$, reflecting the complementarity of quality and output.

**The Firm**

The cost to the firm of producing quantity $x$ at quality $q$ is given by the function $c(x, q; \theta)$, which is parameterized by the cost-state $\theta \in \Theta$, where $\Theta$ is some closed interval of the real number line. The cost of production is $C^2$ and increasing and weakly convex in both quantity and quality and is increasing in $\theta$. Moreover, the cost of output is weakly increasing in quality, $c_{xq} \geq 0$, to reflect the notion that it is more expensive to produce at higher levels of quality. To insure the firm’s problem is

52. The assumption that demand is increasing in $q$ means quantity and quality are complements and is not innocuous. If instead quantity and quality are substitutes then the results of the chapter must change accordingly.

53. A capital letter is used for consumer benefit, and later the quality-adjusted consumer benefit to indicate that it is an aggregate measure of all consumers.

54. Subscripts represent partial derivatives.
well-behaved, $c(x, q)$ is strictly quasi-convex. The regulator cannot observe the firm’s cost, but knows the functional form.

The firm may also value the degree of community benefit generated by its good, reflecting the mission statement of nonprofit institutions. For example, a hospital may prefer to give up some profit in order to provide the community higher quality health care. The firm’s value for community benefit is denoted $\phi(x, q)$ and is concave in its arguments. The firm’s value for community benefit can reflect the regulator’s valuation of consumer benefit, i.e., $\phi(x, q) = B(x, q)$ or, more likely, it simply reflects the values of the firm’s board of directors and is independent of the regulator’s valuation for consumer benefit. The firm’s objective function is $U(\Pi, \phi) = \beta \Pi + (1 - \beta) \phi$, where $\Pi$ are the firm’s profits and $\beta \in [0, 1]$ identifies the relative weight the firm places on profit in relation to community benefit. We further assume that the firm cannot operate with a negative profit reflecting the fact that nonprofit institutions cannot operate with consistent losses so must make up budget shortfalls or face liquidation. It will be shown that this assumption has important ramifications on the regulator’s problem when the firm’s preference for community benefit is sufficiently strong that the firm bumps up against its non-negative profit constraint.

The Regulator

The regulator is a Stackelberg leader and endowed only with the power to establish a unit price $p$ and transfer payment $T$. The transfer payment may come from a fixed

---

55Even a nonprofit hospital, which is not permitted to distribute retained earnings to shareholders, may earn profit that it uses for managerial perks or other expenditures unrelated to conducting business.
payment in a two-part tariff and is assumed to not alter the consumers’ demand. The
prices are enforceable by the regulator.\textsuperscript{56}

The regulator’s objective is to maximize a weighted average of the expected con-
sumer surplus ($CS$) and the firm’s expected profit ($\Pi$). The regulator places a weight
$\alpha \in (\frac{1}{2}, 1]$ on consumer surplus and a weight $1 - \alpha$ on the firm’s profit.\textsuperscript{57}

When the firm possesses superior information about the cost-state, the regulator’s
uncertainty is represented by the distribution $F$ having strictly positive density $f$
over the support $\Theta$. The characteristics of the regulator’s uncertainty are common
knowledge.

### 2.1.1 Quality-Adjusted Cost and Value

Given a payment mechanism $\{p, T\}$, the firm’s objective may be expressed as

$$
\max_q \{U(p, T; \theta) = \beta [px(q, p) - c(x(q, p), q; \theta) + T] + (1 - \beta)\phi(x(q, p)) \} \ \text{s.t.} \ \Pi \geq 0.
$$

Let $q(x, p)$ denote the quality-demand function, that is, $q(x, p)$ denotes the level
of quality required to induce the equilibrium quantity $x$ given the unit price is set to
$p$. Note that the properties of $x(q, p)$ are sufficient to insure the existence of $q(x, p)$.\textsuperscript{58}

It is intuitive to view the firm’s problem as selecting the level of quality that
maximizes its profit for a given price and cost state; however, it is equivalent to view

\textsuperscript{56}For example consumers can report to the regulator any instance in which they were charged a
different price, or were refused service at the regulated price.

\textsuperscript{57}Thus the regulator “cares” more about consumer surplus. Moreover, if $\alpha < 1/2$, then the
regulator’s problem is maximized with unbounded transfers from consumers to the firm.

\textsuperscript{58}More formally, let $D(x, q, p)$ be the implicit function $x - d(q, p)$, where $d(\cdot)$ has the properties
of $x(q, p)$ defined above. Because $d(q, p)$ is continuously differentiable, $D$ has continuous partial
derivative $D_x$, $D_p$, $D_q$, and $D_\theta$ such that $D_x > 0$ and $D_q > 0$ for all $x > 0$, $q > 0$, $p \geq 0$, and
$\theta \in \Theta$. By the Implicit Function Theorem there exists functions $f_1$ and $f_2$ such that $x = f_1(q, p)$
and $q = f_2(x, p)$. 

61
the firm’s problem as choosing the quantity, $x$, which maximizes profit given that it must set the quality, $q(x, p)$, in order to induce an equilibrium demand of $x$. The firm’s objective can therefore be alternatively expressed as

$$\max_x \{ U(p, T; \theta) = \beta [px - c(x, q(p, x); \theta) + T] + (1 - \beta)\phi(x) \} \text{ s.t. } \Pi \geq 0, \quad (2.1)$$

where the firm’s choice variable is now quantity instead of quality.

Similar to Rogerson (1994), define $g(x; p, \theta)$ as the firm’s *quality-adjusted* cost function

$$g(x; p, \theta) = c(x, q(x, p); \theta). \quad (2.2)$$

That is, $g(x; p, \theta)$ denotes the cost of producing the quantity $x$ given that the quality has been adjusted to induce a demand for quantity $x$ when the unit price is $p$ and the cost-state is $\theta$. The relationship between the quality-adjusted marginal cost and the standard marginal cost is

$$\frac{dg}{dx}(x; p, \theta) = \frac{dc}{dx}(x, q(x, p); \theta) = \frac{\partial c}{\partial x}(x, q(x, p); \theta) + \frac{\partial c}{\partial q}(x, q(x, p); \theta) \frac{dq}{dx}(x, p).$$

Thus, the quality-adjusted marginal cost captures both the marginal cost of increasing production, and the marginal cost of increasing the quality necessary to induce the additional demand.

The presence of price in the cost function is unusual, but it allows us to identify a change in cost that occurs as a result of a change in the unit price by the regulator. That is, a change in $g$ with respect to $p$ reflects the change in the firm’s cost that

59 In Rogerson (1994) cost is deterministic so $g(\cdot)$ does not take as arguments the price or demand state.
follows as a consequence of the firm’s reaction to the price response of the consumers:

$$\frac{\partial g}{\partial p}(x; p, \theta) = \frac{dc}{dq}(x, q(x, p), \theta) \frac{dq}{dp}(x, p).$$

It is notable that both $g_x$ and $g_p$ include the term $c_q$ so each only partially captures the change in costs due to a change in the quality level. By using the firm’s quality-adjusted cost function we can more clearly identify the change in cost that occurs because the firm chooses to supply more of the service, $g_x$; and the change in cost that occurs in consequence to a change in the unit price, $g_p$, which is directly controlled by the regulator.

It should be noted that, given the properties of $c(x, q; \theta)$, $g(x; p, \theta)$ must be $C^2$, and strictly increasing and convex in $x$. Moreover, the properties of $c(x, q)$ and $q(x, p)$ further imply $g_{x\theta}(x; p, \theta) \geq 0$; i.e., the marginal cost of output is increasing with the cost state.

Finally, it will also be convenient to define a quality-adjusted consumer benefit function $V$ as:

$$V(x; p) = B(x, q(x, p)). \quad (2.3)$$

As with the quality adjusted cost function, $V(x; p)$ denotes the consumer benefit to consuming the quantity $x$ given that the quality has been adjusted to induce a demand of $x$ when the unit price is $p$. Given the properties of $B$, it must be the case that $V$ is increasing and concave in $x$.

### 2.2 Social Optimum

We start with the case where the regulator and firm have symmetric information regarding all aspects of the model (e.g., the demand, benefit, and cost functions,
the quality, the quantity of output, as well as the cost-state), in order to define the socially optimal outcome. We then proceed to derive the optimal outcome when the regulator and firm have common knowledge of the cost, but the regulator cannot contract directly on output. Although not a pure first-best case, to facilitate the exposition we refer to the solution as first-best when the regulator and firm have symmetric knowledge of the cost state, and we refer to the solution as second-best when the firm has superior knowledge.

The regulator’s objective is to maximize the weighted sum of consumer surplus and profit:

\[ W(p, T; \theta) = \alpha CS(p, T) + (1 - \alpha) \Pi(p, T; \theta). \tag{2.4} \]

Consumer surplus is defined as the gross consumer surplus minus the cost of the good,

\[ CS = V(x; p) - (px + T). \tag{2.5} \]

By substituting \( \Pi \) and \( CS \) into (2.4) and rearranging, the socially optimal outcome is determined by the maximization program:

\[ \max_{x, p, \Pi} V(x; p) - g(x; p, \theta) - \lambda \Pi \]

such that \( \Pi \geq 0 \),

where \( \lambda = (2\alpha - 1)/\alpha > 0 \). By taking the FOCs, the socially optimal outcome is defined as follows.

---

60 This is sometimes referred to as a constrained first-best.

61 A similar optimization program can be derived if the regulator wishes instead to maximize consumer surplus subject to a break-even constraint for the firm.

62 The properties of \( V \) and \( g \) imply the objective function is strictly concave.
Definition 2.2.1. The socially optimal outcome consists of the quantity, \( x^{so} \), and prices, \( \{p^{so}, T^{so}\} \) that equate both the quality-adjusted social marginal benefit of consumption to the quality-adjusted marginal cost and the quality-adjusted marginal benefit of raising the unit price with the quality-adjusted marginal cost, and additionally leaves the firm with zero profit; that is, prices, output, and profit satisfy:

\[
V_x = g_x, \quad (2.6a)
\]
\[
V_p = g_p, \quad (2.6b)
\]
\[
\Pi = 0. \quad (2.6c)
\]

2.3 Regulating with a Marketed Good

Many regulated markets, including those for health services, require that the consumers pay directly for the good or service.\(^6^3\) This helps maintain efficiency and avoid over-consumption. When there are no quality considerations, the price fully determines the quantity demanded; however, because the firm is free to adjust the level of quality, it can manipulate demand, reducing the effectiveness of the regulator’s pricing rule. This section explores how the firm’s ability to manipulate demand affects what the regulator can achieve.

\(^6^3\)In markets for health care consumers are generally insured to some degree, but higher medical care costs will invariably lead to higher health insurance premiums so \( x \) can be thought of as a crude reduced-form measure of this demand response to health care costs through the insurance channel.
2.3.1 Symmetric Information about $\theta$

When the regulator cannot contract on output, then as a Stackelberg leader, it must offer a contract $\{p, T\}$ which maximizes its objective given how the firm will respond to the payment rule. Denote $x^*$ as the quantity maximizing the firm’s objective subject to a nonnegative profit constraint. The first order condition from the firm’s problem, (2.1), shows that the firm’s maximizer $x^*$ solves one of two possible equations. First, if the firm’s maximization constraint is not binding (i.e., the Lagrange multiplier is zero) we refer to the firm as having a mixed objective and $x^*$ solves

$$p + \frac{(1-\beta)}{\beta} \phi'(x^*(p, \theta)) = g_x(x^*(p, \theta); p, \theta).$$

(2.7)

On the other hand, when the firm’s constraint binds then it is a pure output-maximizer and $x^*$ solves

$$px^*(p, \theta) + T = g(x^*(p, \theta); p, \theta).$$

(2.8)

Given the firm’s best response function $x^*$, the regulator’s problem (RP-M) can be expressed as

$$\max_{p, \Pi} V(x^*(p, \theta); p) - g(x^*(p, \theta); p, \theta) - \lambda \Pi \text{ subject to } \Pi \geq 0.$$

We have not yet established that the regulator’s problem is concave; i.e. $D^2W(p, \Pi) < 0$. To insure concavity in $p$ we must assume the bordered Hessian $|\mathcal{H}|$ is positive definite. Because the regulator’s problem can be expressed equivalently as

$$\max_{x, p, \Pi} V(x; p, \theta) - g(x; p, \theta) \text{ s.t. } \Pi \geq 0 \text{ and } x = x^*(p, \theta).$$

The relevant bordered Hessian (See the regulator’s problem linearly so is ignored) is thus defined as

$$|\mathcal{H}| = \begin{bmatrix} 0 & 1 & -dx^*/dp \\ 1 & V_{xx} - g_{xx} & V_{xp} - g_{xp} \\ -dx^*/dp & V_{xp} - g_{xp} & V_{pp} - g_{pp} + (V_x - g_x)d^2x^*/dp^2 \end{bmatrix} > 0.$$
The first order condition with respect to \( p \) identifies the first-best price \( p^{fb} \) as the \( p \geq 0 \) solving
\[
V_x(x^*(p^{fb}, \theta); p^{fb}) \frac{dx^*}{dp}(p^{fb}, \theta) + V_p(x^*(p^{fb}, \theta); p^{fb}) = \\
g_x(x^*(p^{fb}, \theta); p^{fb}, \theta) \frac{dx^*}{dp}(p^{fb}, \theta) + g_p(x^*(p^{fb}, \theta); p^{fb}, \theta).
\]
(2.9)

The first-best price consists of the price equating the marginal benefit of increasing the unit price to the marginal cost.

To ease the interpretation of (2.9), we start by interpreting the marginal cost term. The change in the firm’s cost can be thought of as a Slutsky-like decomposition of the change in cost with respect to a price change. The first term on the RHS, \( g_x(dx^*/dp) \), identifies the change in the firm’s cost that results from adjusting output to take advantage of the change in revenue following an increase in price and should be thought of as a \textit{revenue effect}. The second term on the RHS, \( g_p \), identifies the change in the firm’s cost associated with raising quality in order to maintain the same quantity of output following a price increase and should be interpreted as a \textit{demand effect} from an increase in price. The marginal benefit consists of the social benefit derived from the firm’s adjustment in the equilibrium output, \( V_x(dx^*/dp) \), and the benefit gained (or lost) due to the demand effect from a change in unit price.

If a regulatory policy is to achieve the first-best outcome, it must meet two conditions: (i), the firm must be held to zero profits; and (ii), the firm must be induced to produce the efficient quantity \( x^{fb} \). By recognizing that for any \( p \) the firm will choose the output equating its marginal benefit of output to its quality-adjusted marginal cost of production the first-best payment policy can be derived. The following proposition reports the optimal payment policy.
Proposition 7. The optimal payment rule with symmetric cost information for a mixed-objective firm consists of the unique unit price $p^{fb}(\theta)$ and transfer payment $T^{fb}(\theta)$ satisfying:

$$
p^{fb}(\theta) = V_x(x^{fb};p^{fb}) + \frac{V_p(x^{fb};p^{fb}(\theta)) - g_p(x^{fb};p^{fb}(\theta),\theta)}{dx^*(p^{fb}(\theta),\theta)/dp} - \frac{(1-\beta)}{\beta} \phi'(x^{fb}(\theta)),
$$

$$
T^{fb}(\theta) = g(x^{fb};p^{fb}(\theta),\theta) - p^{fb}(\theta)x^{fb},
$$

and for a pure output-maximizing firm

$$
p^{fb}(\theta) = p^{so}(\theta),
$$

$$
T^{fb}(\theta) = g(x^{fb};p^{fb},\theta) - p^{fb}x^{fb},
$$

where $p(\theta) \leq g_x(x^{fb}(\theta);p^{fb}(\theta),\theta)$ for all $\theta \in \Theta$.

The intuition for the unit payments are as follows. For a mixed-objective firm, the first two terms for $p^{fb}(\theta)$ account for the revenue and demand effects of adjusting price. Given a unit price, the firm will adjust the quality to insure that its quality-adjusted marginal cost of production equals its marginal benefit; i.e., the unit price. Thus the first term in $p^{fb}(\theta)$ is the marginal social value of changing the equilibrium quantity. The firm adjusts the quality level to compensate for the demand response to the change in price so the second term in $p^{fb}$ is present to account for the change in social value due to the demand effect of a change in price. The firm is rewarded for each unit sold and not for the quality level of the good, therefore it must be compensated for each unit of additional surplus and the total change in social surplus due to the demand effect, $(V_p - g_p)$, is divided by the change in the equilibrium quantity, $dx^*/dp$. The third term adjusts the price to account for the firm’s preference for community
benefit. The more weight the firm places on community benefit, the less the regulator must compensate the firm to induce the first-best outcome. Lastly, when the firm is a pure output maximizer then the regulator can simply equate the unit price to the socially optimal unit price and adjust the transfer to induce the firm to produce the socially optimal level of output.

It is clear from the definition of $p^{fb}(\theta)$ in Proposition 7, that if the marginal benefit and marginal cost of a price change are equivalent at $x^{fb}$ (i.e., $V_p(x^{fb}; p^{fb}) = g_p(x^{fb}; p^{fb})$), then the first-best and socially optimal unit prices must also be equivalent. Because the firm chooses $x$ so that $U_x = 0$, that is, because the firm maximizes based on a marginal consumer’s valuation of quality, the level of quality will differ from the social optimum. For example, quality will be undersupplied if consumer benefit is the area under the demand curve, or it may be oversupplied if there are negative externalities to consumption. The following proposition formally reports this result.

**Proposition 8.** Given a price, $p$, output (and quality) may be over- or undersupplied relative to the socially-preferred level. The output differs from the socially preferred

---

65This is the analog to a result first reported by Spence (1975) and Baron (1981), who studied the provision of quality in a more general framework.

66In Baron (1981) the regulator’s value function is simply the area under the demand curve; i.e., $B = \int_0^x P(\bar{x}, q) d\bar{x}$ where $P$ is the inverse demand function. Thus, without externalities, $V_x = p + \int_0^x P_q(\bar{x}, q) q_x d\bar{x} \neq p$ and $V_x > p^{fb} = g_x$ implying both output and quality are undersupplied. If the regulator’s measure of consumer benefit accounts for some negative externality then it may be the case that $V_x < p$. As an example consider the social surplus function characterized as a downward, parallel shift of the demand curve: $S = \int_0^x (P(\bar{x}, q) - \beta) d\bar{x}$. The FOC yields, $V_x = p - \beta + \int_0^x P_q(\bar{x}, q) q_x d\bar{x}$. Thus, $V_x < p$ if $\beta > \int_0^x P_q(\bar{x}, q) q_x d\bar{x}$; that is, the quality adjusted marginal surplus is less than the unit price if the cost of the negative externality exceeds the marginal benefit of the change in consumption caused by a change in the quality.
output according to the rule:

\[ x^* \overset{\text{arg max}_{x}}{\succeq} V - g \quad \text{when} \quad p \overset{\text{arg max}_{x}}{\succeq} V_x(x^*; p) - \frac{(1-\beta)}{\beta} \phi'. \]

Proof. See Appendix B.

The reason the regulator is unable to induce the social optimum with symmetric information is because the firm’s choice of quality remains non-contractible. In effect, the regulator must use the single instrument of the unit price to control the firm’s choice in quality while simultaneously adjusting consumer demand to achieve a socially optimal outcome.

2.3.2 Asymmetric Information about \( \theta \)

Under asymmetric information the problem is a standard adverse selection screening problem, thus to insure there exists a separating equilibrium we require type separation across cost-states. As is common in screening problems, we impose the single crossing property on the firm’s value function.

Definition 2.3.1. The single-crossing property holds if the firm’s marginal rate of substitution (MRS) of price for transfer payment \( (U_p/U_T) \) is monotonic in \( \theta \) for all \( \theta \in \Theta \).

Without loss of generality, we will take advantage of the revelation principle and restrict the analysis to truthful direct mechanisms (Dasgupta et al., 1979; Myerson, 1979). In a direct revelation mechanism the firm announces the state parameter which optimizes its state-dependent value function \( \mathcal{U} \). Because the firm’s objective
is to maximize a weighted sum of profit and community benefit, its state-dependent
value function is defined as

\[ U(\hat{\theta}, \theta) = \beta [p(\hat{\theta})x^*(p(\hat{\theta}), \theta) - g(x^*(p(\hat{\theta}), \theta); p(\hat{\theta}), \theta) + T(\hat{\theta})] + (1 - \beta) \phi(x^*(p(\hat{\theta}), \theta)), \]

where \( \theta \) is the true state and \( \hat{\theta} \) is the firm’s announcement. The regulator’s objec-
tive is to maximize the total expected social surplus subject to standard individual
rationality and incentive compatibility constraints. The regulator’s problem may be
expressed as

\[
\max_{p(\theta), U(\theta)} \int_{\Theta} \left\{ V(x^*(p(\theta), \theta); p(\theta)) - g(x^*(p(\theta), \theta); p(\theta), \theta) - \frac{1}{\beta} (U(\theta) - (1 - \beta) \phi(\theta)) \right\} dF(\theta),
\]

subject to

\[
U(\theta) \geq 0 \quad \forall \theta \in \Theta \quad \text{(Individual Rationality)}
\]
\[
U(\theta, \theta) \geq U(\hat{\theta}, \theta) \quad \forall \hat{\theta}, \theta \in \Theta \quad \text{(Incentive Compatibility)}.
\]

where \( U(\theta) = U(\theta, \theta) \) and \( \phi(\theta) = \phi(x^*(p(\theta), \theta)) \). Note that the firm’s rents are
expressed as the difference between the firm’s total value for output, \( U(\theta) \), and the
firm’s value for community benefit given a truthful report of the cost state, \( \phi(\theta) \). The
expression \( U(\theta) - (1 - \beta) \phi(\theta) \) is divided by \( \beta \) to account for the fact that the firm
places weight \( \beta \) on profit whereas the regulator is concerned about the social loss of
any profit.

The following lemma characterizes the necessary and sufficient conditions for an
incentive compatible payment policy.

**Lemma 4.** *The menu of two-part tariffs \{p(\theta), T(\theta)\}_{\theta \in \Theta} is incentive compatible if and only if it satisfies the conditions*
\[
\frac{dU(\theta)}{d\theta} = \frac{\partial U(\theta)}{\partial \theta} = \begin{cases} -\beta g_\theta(x^*(p(\theta), \theta); p(\theta), \theta) \\ -(1 - \beta) \phi'(x) \frac{g_\theta(x^*(p, \theta))}{g_x(x^*(p, \theta) - p) - p} \end{cases}
\]
when \(\Pi(x^*) > 0\), otherwise.

(ii) \(\text{sign}[dp/d\theta] = \text{sign} \left[ \frac{\partial}{\partial \theta} \left( \frac{U_p}{U_T} \right) \right] \).

Proof. See Appendix B. \(\square\)

Condition (i) is critical to the results of this chapter and is a consequence of the fact the firm will lower the service quality in higher cost-states unless it receives a higher unit payment to compensate for higher marginal costs. Providing a higher unit payment in higher states generates an incentive for the firm to misreport the state as higher than it is. Consequently incentive compatibility requires the firm receive higher rents in lower cost states to counter the desire to misreport the cost state as being high. When the firm’s optimal choice leaves it with some positive profit, then anything that increases the rate at which the firm’s costs increase with the state will cause it to receive higher rents in all states. When the firm does not earn positive profit then it chooses the output that leaves it with zero profit. Again, the firm has an incentive to misreport the state if doing so allows it to increase output. The fraction \(g_\theta/(g_x - p)\) identifies the trajectory for the firm’s choice of output across cost states, which notably is a direct function of the unit price. Both conditions are negative, thus the firm’s informational advantage is necessarily decreasing with the cost state.

Condition (ii) is specific to the use of the two-part tariff and reports that any payment policy satisfying incentive compatibility must offer a unit payment that moves in the same direction as the firm’s MRS of price for the fixed transfer. Because the firm’s MRS of price for fixed transfer must be monotone for type separation, it
follows that the regulator’s pricing rule is also monotone; however, it is not restricted to being either an increasing or a decreasing function of the state parameter.

Using Lemma 4 we can also identify the firm’s informational rents when its optimal choice of output leaves it with positive profit. The following corollary reports this result.

**Corollary 2.** An incentive compatible menu of two-part tariffs \( \{p(\theta), T(\theta)\}_{\theta \in \Theta} \) must leave a marginal optimizer with expected information rents:

\[
E\Pi(\theta) = \int_{\Theta} \left\{ \frac{1-\beta}{\beta} \left[ \phi(\theta) - \phi(x^*(\theta)) \right] + \frac{F(\theta)}{F(\theta)} g_\theta(x^*(\theta); p, \theta) \right\} dF(\theta).
\] (2.11)

*Proof.* See Appendix B.

When the firm’s optimal choice of output leaves it with some positive profit, the firm acts in a similar manner to a purely profit maximizing firm in that it increases output until the marginal benefit of increasing output equals the marginal cost. The more weight the firm places on community benefit, the more it helps the regulator by limiting its own rents and when the firm is purely output-maximizing, its behavior, and hence the regulator’s problem, changes substantially.

If the firm’s preference for community benefit causes the non-negative profit constraint to bind, then the regulator cannot focus on the firm’s rents in any cost state, but must instead insure that the firm does not misreport the cost state in order to over-produce for the sake of output. As a consequence, the approach to solving the optimal contract differs depending on if the firm has a mixed objective, or is purely output-maximizing. The next two subsections analyzes the optimal contract for a firm with mixed objectives, and a purely output-maximizing firm, respectively.
A Mixed-Objectives Firm

We will solve the regulator’s problem using optimal control. The regulator’s problem is to maximize Eq. (2.10) subject to the constraints

\[ \frac{dU}{d\theta} = -\beta g(\theta^*(p(\theta), \theta); p(\theta), \theta), \]

\[ U(\theta) = \varphi(\theta). \]

The Hamiltonian for the regulator’s problem can thus be expressed as

\[ H = \left\{ V(x^*(p(\theta), \theta); p(\theta)) - g(x^*(p(\theta), \theta); p(\theta), \theta) - \frac{1}{\beta} (U(\theta) - (1 - \beta)\varphi(\theta)) \right\} f(\theta) \]

\[ - \delta(\theta)\beta g(\theta^*(p(\theta), \theta); p(\theta), \theta), \]

where \( U(\theta) \) is the state variable, \( p(\theta) \) is the control, and \( \delta(\theta) \) is the Pontryagin multiplier. By the maximum principle \( dH/dU = -d\delta/d\theta = -(\lambda/\beta)f(\theta) \). The boundary condition at \( \theta \) is unconstrained so the transversality condition at \( \theta = \theta_0 \) is \( \delta(\theta) = 0 \). Integrating \( d\delta/d\theta \) yields \( \delta(\theta) = (\lambda/\beta)F(\theta) \).

Plugging in \( \delta(\theta) \) and taking the first-order condition of the Hamiltonian yields

\[ \frac{d}{dp} \left\{ V(x^*(p(\theta), \theta); p(\theta)) - g(x^*(p(\theta), \theta); p(\theta), \theta) \right\} = \lambda \left[ \frac{F(\theta)}{f(\theta)} \frac{d}{dp} \left\{ g(\theta^*(x^*; p, \theta) \right\} - \frac{1 - \beta}{\beta} \phi'(x^*) \frac{dx^*}{dp} \right] \]

(2.12)

The price \( p^{sb} \) solving Eq. (2.12) is the second-best price given the regulator’s constraints and (2.12) equates the marginal change in social surplus to the marginal social loss. Eq. (2.12) has a familiar interpretation. Increasing the payment to the firm in cost-states \([\theta, \theta + d\theta]\), which number \( f(\theta)d\theta \), by \( dp \) increases the social surplus by \( \left[ V_p(x^*; p, \theta) - g_p(x^*; p, \theta) + (V_z(x^*; p, \theta) - g_z(x^*; p, \theta)) \frac{dx^*}{dp} \right] dp \). From (i) of Lemma 4, the increase in output simultaneously increases the firm’s rent in
cost-states \([\theta, \theta]\), which number \(F(\theta)\), by \(\frac{d}{dp}\{g_\theta(x^*; p, \theta)\}\)dp, offset by the degree to which the firm’s value for community benefit changes with respect to the increase in price, \(\frac{1-\beta}{\beta} \phi'(x^*) \frac{dx^*}{dp} f(\theta)\)dp. The total social cost of the increase in the firm’s rent is \(\lambda F(\theta) \frac{d}{dp}\{g_\theta(x^*; p, \theta) - \frac{1-\beta}{\beta} \phi'(x^*) \frac{dx^*}{dp} f(\theta)\}\)dp.

If the price \(p_{sb}\) solving Eq. (2.12) does not satisfy condition (ii) of Lemma 4 then the payment rule is not incentive compatible and the regulator is unable to extract any usable information from the firm. Consequently, permitting the firm to use its private knowledge to select the unit price is too socially costly. In such a case the regulator cannot allow the firm to choose the price and must set a fixed-price, \(p(\theta) = \hat{p} \in [p_{sb}^{(\theta)}, p_{sb}^{(\theta)}]\).\(^{67}\) Caillaud et al. (1988) and Laffont and Tirole (1993, pg. 161) refer to this as “the phenomenon of nonresponsiveness of the allocation with respect to private information.”

Identifying if condition (ii) of Lemma 4 is satisfied requires either establishing functional forms for the cost, value, and demand functions or making several strict assumptions on the higher order derivatives of these functions.\(^{68,69}\) To maintain generality and avoid making assumptions on higher-order derivatives that do not have an economic justification, for the remainder of this section we will assume the properties of the cost, value, and demand functions are sufficient to insure that the

\(^{67}\)This represents an analog to the case of decreasing marginal costs in Lewis and Sappington (1988a) and the case of an increasing labor allocation in a self-managed firm in Guesnerie and Laffont (1984). In both cases the optimal regulatory policy fails incentive compatibility eliminating the regulator’s ability to extract any information about the state of the world.

\(^{68}\)For example, the signs for \(g_{\theta p}, g_{\theta px}, g_{\theta xx}\) and \(g_{\theta x}\) must be established.

\(^{69}\)See Rogerson (1987) for a discussion on the necessary properties of the model’s primitives that allow for an implementable payment policy for a similar principal-agent problem.
relationship reported in Lemma 4 is satisfied. When the condition is not satisfied the optimal policy will require some degree of pooling.

Returning to the first-order condition of the regulator’s problem given by Eq. (2.12) we can analyze how the firm’s informational advantage distort the prices and output. Recall that the first-best solution is the contract that sets the left-hand side of (2.12) equal to zero. Because the firm’s information rent may be either increasing or decreasing with the unit payment, however, the right-hand side of (2.12) may be either less than or greater than zero, resulting in either \( p^{sb} > p^{fb} \) or \( p^{sb} < p^{fb} \), respectively. Intuitively, the regulator will shade the price up or down from the first-best price to limit the information rents attained by the firm. That is, if the point-wise derivative of the firm’s rents with respect to price is increasing with the unit payment \( (dE/\partial p > 0) \), then the regulator will decrease the payment to limit the firm’s rent, and vice-versa. The following proposition formally reports this relationship between first-and second-best unit payments.

**Proposition 9.** The relative magnitude of the second- to the first-best unit price is inversely related to the effect a price change has on the firm’s information rent:

\[
P^{sb} = \begin{cases} 
  p^{fb} & \text{when } \frac{E(\theta)}{f(\theta)} d\{g_\theta(x^*; p, \theta)\} / dp > \frac{1-\beta}{\beta} \phi'(x^*) \frac{dx^*}{dp}, \\
  p^{fb} & \text{when } \frac{E(\theta)}{f(\theta)} d\{g_\theta(x^*; p, \theta)\} / dp = \frac{1-\beta}{\beta} \phi'(x^*) \frac{dx^*}{dp}, \\
  > p^{fb} & \text{when } \frac{E(\theta)}{f(\theta)} d\{g_\theta(x^*; p, \theta)\} / dp < \frac{1-\beta}{\beta} \phi'(x^*) \frac{dx^*}{dp}.
\end{cases}
\]

**Proof.** See Appendix B.

More importantly, it follows that if the second-best unit price may be higher or lower than the first-best, then the second-best output (hence the quality) may be under- or over-supplied relative to the first-best as well. For example, when the
second-best price is less than the first-best, and the firm’s optimal choice of output is increasing in the price \((dx^*/dp > 0)\), then the second-best level of output will necessarily be below first-best and when the firm’s optimal choice of output is decreasing in the price \((dx^*/dp < 0)\), then the second-best level of output will exceed the first-best output. When the second-best price is greater than the first-best the results are of course reversed.

It comes as no surprise that by adding \(\phi(x)\) as a general expression of the firm’s valuation of community benefit we can achieve an ambiguous result. For example, if a hospital’s board of directors values community benefit substantially more than the regulator then the regulator may have no choice but to allow the hospital to over-supply hospital services to some degree in order to maintain incentive compatibility. It is surprising to note, however, that this outcome is not necessarily dependent on the characteristics of \(\phi\); that is, even with a pure profit-maximizing firm, the second-best outcome may be under- or over-supplied relative to first-best. The direction of the distortion is determined by how the firm’s choice of output and rents change with a change in the unit payment. To see this, we start with the definition of the firm’s informational rents identified in corollary 2. When the firm is a pure profit-maximizer, then \(\beta = 1\) and taking the point-wise derivative of the firm’s expected rents yields:

\[
\frac{d}{dp} \left\{ \frac{F(\theta)}{f(\theta)} \frac{\partial g(x^*; p, \theta)}{\partial \theta} \right\} = \frac{F(\theta)}{f(\theta)} \left[ \frac{\partial^2 g}{\partial \theta \partial p} + \frac{\partial^2 g}{\partial \theta \partial x} \frac{dx^*}{dp} \right].
\]

The first term within the brackets on the RHS, \(g_{\theta p}\), identifies the direct change to the firm’s rent that follows from an increase in the unit price. The partial change to the firm’s rents following a price change (in all cost states) is the change to the firm’s rent that comes about from adjusting quality to maintain the same equilibrium.
output. Increasing quality increases the cost of production, which in turn increases the incentive to misreport a high cost-state. Therefore the firm’s rents must increase in all states in the price dimension, \( g_{\theta p} \geq 0 \). The second term \( g_{\theta x} \frac{dx^*/dp}{g_{\theta x}} \) identifies the indirect change in the firm’s rents that follows from adjusting the equilibrium quantity demanded due to a price change. Increasing output increases the cost of production therefore the firm’s rents increase in all states along the quantity dimension as well.

Therefore, \( d\{g_{\theta}\}/dp < 0 \) if and only if the firm’s best response to a price increase is to decrease the equilibrium quantity sufficiently as to lower its overall costs in every state \( \theta \). That is, the firm’s rents are decreasing with a price increase if and only if \( \frac{dx^*/dp}{g_{\theta p}/g_{\theta x}} < 0 \).

When \( dx^*/dp > 0 \) the second-best output is undersupplied relative to first-best because the firm’s rents are unambiguously increasing with the unit payment. The regulator thus sets a price below the first-best price in order to limit the firm’s rents, resulting in the undersupply. On the other hand, when \( dx^*/dp < 0 \) the outcome depends on whether or not \( dx^*/dp \) is sufficiently negative to flip the firm’s rents so that they are decreasing in the unit payment. If it is, then the regulator will have to set a price above the first-best price to limit the firm’s rents. Because the firm’s optimal choice of output varies inversely with the price, the higher unit payment causes the firm to still choose an output below the first-best. Furthermore, if \( dx^*/dp \) is insufficiently negative, then the firm’s rents are still increasing with the price and the regulator will again choose a price below first-best. Because the firm’s best response
to a decrease in price is to *increase* output, this results in an oversupply relative to first-best. The following proposition formally identifies these three cases.\(^{70}\)

**Proposition 10.** With asymmetric cost information and a pure profit-maximizing firm, the relative size of the second- to first-best price and output is determined by the rules:

\[(i) \quad \frac{dx^*}{dp} < -\frac{g_{\theta p}}{g_{\theta x}} < 0 \Rightarrow \frac{d}{dp} \left[ \frac{\partial \Pi}{\partial \theta} \right] < 0 \Rightarrow \begin{cases} p^{sb}(\theta) > p^{fb}(\theta) \\ x^{sb}(\theta) < x^{fb}(\theta) \end{cases} \]

\[(ii) \quad -\frac{g_{\theta p}}{g_{\theta x}} < \frac{dx^*}{dp} < 0 \Rightarrow \frac{d}{dp} \left[ \frac{\partial \Pi}{\partial \theta} \right] > 0 \Rightarrow \begin{cases} p^{sb}(\theta) < p^{fb}(\theta) \\ x^{sb}(\theta) > x^{fb}(\theta) \end{cases} \]

\[(iii) \quad -\frac{g_{\theta p}}{g_{\theta x}} < 0 < \frac{dx^*}{dp} \Rightarrow \frac{d}{dp} \left[ \frac{\partial \Pi}{\partial \theta} \right] > 0 \Rightarrow \begin{cases} p^{sb}(\theta) < p^{fb}(\theta) \\ x^{sb}(\theta) < x^{fb}(\theta) \end{cases} \]

for any \(\theta \in (\widehat{\theta}, \overline{\theta})\) and at \(\overline{\theta}\), \(p^{sb}(\overline{\theta}) = p^{fb}(\overline{\theta})\) and \(x^{sb}(\overline{\theta}) = x^{fb}(\overline{\theta})\).

**Proof.** The proof follows immediately from Proposition 9 and from identifying \(\text{sign}[d\{g_0\}/dp] \). \(\square\)

Figure 2.1 graphically represents the three cases identified by Proposition 10. The horizontal axis represents the quantity of output, and the vertical axis the price in dollars. The level curves for a representative \(g_\theta(x^*; p, \theta')\) are displayed and should be thought of as iso-rent curves since the firm’s informational rents are a function of how its costs change with the state parameter. The iso-rent curves are increasing away from the origin; i.e., \(g_{\theta p} > 0\) and \(g_{\theta x} > 0\). In Figures 2.1(a) and 2.1(b) the firm’s optimizer decreases with an increase in price and in figure 2.1(c) it increases. Figure 2.1(a) represents case \((i)\) of proposition 10 as the decrease in output is sufficient to

\(^{70}\)Identification of the output distortion requires one more technical assumption that \(\text{sign}[dx^*(p^{sb})/dp] = \text{sign}[dx^*(p^{fb})/dp]\); i.e., that the distortions are not dramatic enough that the firm’s output response to a price change moves in opposite directions.
drop to a lower level curve resulting in a decrease in the firm’s rents. In Figure 2.1(b) the decrease in output is insufficient. Representing case (ii), the decrease in output results in a jump up to a higher level curve and an increase in rents. Figure 2.1(c) corresponds with case (iii) as the change in output following a price increase also results in a straightforward jump to a higher level curve indicating an increase in the firm’s rents. It should be noted that movement across iso-rent curves need not be monotone across cost states when the firm’s optimal choice of output is decreasing with price. Thus, for some functional forms, in some states the regulator’s problem will satisfy condition (i) and for others condition (ii) of Proposition 10.

Figure 2.1: Graphical Examples of the Pricing Rules from Proposition 10

Proposition 10 informs us that we cannot predict a priori, without having specific functional forms for the cost, demand, and value functions, how the firm’s informational advantage will distort output when quality is added to the model, even for a traditional profit-maximizing firm. This qualitative result is in sharp contrast to the
results found in the variety of research cited in the introduction where the firm’s informational advantage results in a strictly downward distortion from first-best. With its choice of quality, the firm is able to adjust the quantity demanded for a given unit price. In consequence, the unit price does not just alter the quantity demanded by consumers, but it alters the quality level the firm chooses, which in turn also alters the quantity demanded. The two sources of demand adjustment may lead to either an over- or under-supply relative to first best. Identifying the distortion to quality is more problematic. For example, in case (iii) it is clear that if prices and output are below first-best levels, then quality must be undersupplied, but in cases (i) and (ii) quality may be under- or oversupplied relative to first-best. When the information asymmetry is in demand, however, we show in Proposition 16 that quality may be unambiguously oversupplied.

An Output-Maximizing Firm

We will again use optimal control to analyze the regulator’s problem; however, we must choose a different state variable because the firm solves half of the regulator’s problem by leaving itself with zero profit and the regulator does not care about the value the firm puts on output.

Because the regulator cares about maximizing social surplus let $S(\theta) \equiv V(x; p, \theta) - g(x; p, \theta)$, be the state variable and let $p(\theta) = p^{np}(\theta)$ be the control variable. The regulator’s objective is to maximize $\int_{\Theta} S(\theta) dF(\theta)$ subject to

$$\frac{dS}{d\theta} = \frac{d}{d\theta} \{V - g\} = (V_x - g_x) \frac{dx}{d\theta} + (V_p - g_p + (V_x - g_x) \frac{dx}{dp}) \frac{dp^{np}}{d\theta} - g_\theta.$$  (2.13)
Let $\delta(\theta)$ be the Pontryagin multiplier, then by the maximum principle $dH/dS = -d\delta/d\theta = -f(\theta)$. The boundary condition at $\theta$ is unconstrained so the transversality condition at $\theta = \tilde{\theta}$ is $\delta(\tilde{\theta}) = 0$. Thus, integrating $d\delta/d\theta$ yields $\delta(\theta) = F(\theta)$. The first-order condition of the Hamiltonian yields:

$$\frac{d}{dp}\left\{(V_x - g_x)\frac{dx^*}{d\theta} + (V_p - g_p + (V_x - g_x)\frac{dx^*}{dp})\frac{dp^{np}}{d\theta} - g_\theta\right\} = 0, \quad (2.14)$$

where $p^{np}$ is the unit price solving the first-order condition. When $dx^*/dp = dx^{fb}/dp$ and $dp^{np}/d\theta = dp^{fb}/d\theta$ then the expression on the LHS of (2.14) is, by definition, equal to zero. Unfortunately there does not exist a closed form solution to the regulator’s problem and it must be solved via numerical methods.\footnote{For this reason we cannot identify how the second-best price is distorted away from the first-best price beyond demonstrating that they are not equal.}

It can be shown, however, that the regulator will not generically be able to achieve the first-best with an output-maximizing firm. Recall that Lemma 4 indicates that $dx^*/d\theta = -g_\theta(x^*;p,\theta)/(g_x(x^*;p,\theta) - p)$. Thus the trajectory of $x^*$ across cost-states is a direct function of the unit payment. Lemma 8 (in the appendix) reports that $p \leq g_x$ for an output maximizing firm, therefore $dx^*/d\theta < 0$ and, using the integrable form of the envelope theorem (Milgrom, 2001, pg. 67), $x^*(\theta) = x^*(\tilde{\theta}(\theta),\theta)$ may be expressed as

$$x^*(\tilde{\theta}(\theta),\theta) = x^*(\tilde{\theta}(\bar{\theta}),\bar{\theta}) - \int_\theta^{\bar{\theta}} \frac{\partial x^*}{\partial \theta}(\tilde{\theta}(\tilde{\theta}),\tilde{\theta}) d\tilde{\theta},$$

where $\partial x^*/\partial \theta = -g_\theta(g_x - p)$ as established by Lemma 4.

Thus, if the regulator can design a contract so that $x^*(\bar{\theta}) = x^{fb}(\bar{\theta})$ and $dx^*/d\theta = dx^{fb}/d\theta$ in all $\theta \in \Theta$ then it will induce the first-best output in all $\theta$. The following
proposition identifies the unique payment policy and accompanying necessary and sufficient conditions that accomplishes this outcome.\textsuperscript{72}

**Proposition 11.** The menu of two-part tariffs \( \{p^{np}(\theta), T^{np}(\theta)\}_{\theta \in \Theta} \) induces an output-maximizing firm to produce the first-best quantity \( x^{fb} \), where

\[
\begin{align*}
p^{np}(\theta) &= g_x(x^{fb}(\theta); p^{np}(\theta), \theta) + \frac{g_\theta(x^{fb}(\theta); p^{np}(\theta), \theta)}{dx^{fb}(\theta)/d\theta}, \\
T^{np}(\theta) &= g(x^{fb}(\theta); p^{np}(\theta), \theta) - p^{np}(\theta)x^{fb}(\theta),
\end{align*}
\]

if and only if \( dx^{fb}/d\theta < 0, dp^{np}/d\theta < 0, \) and \( p^{np}(\theta) \geq 0 \) for all \( \theta \in \Theta \).

**Proof.** See Appendix B. \qed

The payment rule reported by Proposition 11 is sufficient to induce the first-best output; however, it does not induce the first-best outcome. This follows because the equilibrium quantity is determined by both the price consumers pay and the level of quality established by the firm. Because the regulator is constrained to use the unit payment to control the firm’s choice of output, \( x^* \), it cannot also set the price to the first-best price. Given the concavity of the regulator’s problem the first-best unit price is unique as is the payment policy inducing the first-best output, and they are not generically equivalent. Put another way, the firm has one instrument, with which to solve two equations— and it cannot. The following proposition formally states this result.

\textsuperscript{72}Existence of a \( p^{np} \) that solves the payment rule in Proposition 11 is not guaranteed for all quality adjusted cost functions, \( g \). For example, if \( g_\theta/(dx^{fb}/d\theta) \) is too negative then there may not exist a \( p^{np}(\theta) > 0 \) that induces the first-best output in all \( \theta \) when the good is marketed. There always exists a \( p^{np}(\theta) \) when the good is nonmarketed, however, because \( p^{np}(\theta) \) can be negative.
**Proposition 12.** The regulator cannot generically induce the first-best outcome for a pure output-maximizing firm.

*Proof.* See Appendix B.

Note that because the payment policy identified by Proposition 11 does not induce the first-best outcome with a marketed good, it is also not the second-best payment rule when the good is marketed.

### 2.4 A Nonmarketed Good

When the good or service is not marketed then there is no direct demand-response to the contract and the regulator only needs to account for how the firm best-responds to the payment rule it sets.

#### 2.4.1 Symmetric Information about $\theta$

The regulator’s problem continues to be that of designing a menu of two-part tariffs $\{p(\theta), T(\theta)\}_{\theta \in \Theta}$ which maximizes a weighted sum of consumer surplus and profit given the firm will choose the quantity solving (2.8). Because payment is made directly by the regulator using public funds raised through taxation, we introduce a shadow cost to public funds, $\gamma > 0$. In this way, every $1$ paid by the regulator for the firm’s good has a total social cost of $(1 + \gamma)$. Net consumer surplus is now defined as:

$$CS = B(x, q) - (1 + \gamma)(px + T).$$

As with a marketed good, it is convenient to define and work with the quality-adjusted cost and consumer value functions. Consumer demand is unaffected by the unit price,
though, so the quality-adjusted functions are no longer take price as an argument; i.e.,

\[ g(x; \theta) = c(x, q(x); \theta), \]
\[ V(x) = B(x, q(x)). \]

The regulator’s problem (RP-NM) with symmetric cost information can be expressed as

\[
\max_{p, \Pi} V(x^*(p, \theta)) - (1 + \gamma)g(x^*(p, \theta); \theta) - (\lambda + \gamma)\Pi \quad \text{subject to } \Pi \geq 0
\]

The firm’s profit still enters the regulator’s problem negatively; however, the shadow-cost to public funds increases the loss to social welfare that positive firm profit generates. Nevertheless, removing the unit price from consumer demand simplifies the regulator’s problem. For example, because the objective function is strictly concave in \( x \), it is strictly concave in \( p \) without any further assumptions.\(^{73}\) Moreover, the firm’s best response to a price increase is to increase output.\(^{74}\) This follows because an increase in the unit price increases the firm’s revenue with no concomitant increase in cost, therefore the firm will increase output until its marginal cost of production again equals the higher unit payment.

The first order condition of the regulator’s problem (RP-NM) yields the first-best payment rule with symmetric information.

\(^{73}\)To see why, start with \( d^2\{V - g\}/dp^2 = (V_{xx} - g_{xx})\frac{dx^*}{dp} \). The term in parenthesis is strictly negative thus concavity only requires \( dx^*/dp > 0 \). Whereas \( dx^*/dp \) could be either positive or negative when demand is a function of price, it is strictly positive for a nonmarketed good.

\(^{74}\)From the conjugate pairs theorem \( sign[dx^*/dp] = sign[\Pi_{xp}] \) and \( \Pi_{xp} = 1 \).
Proposition 13. The optimal payment rule for a nonmarketed good with symmetric cost and demand information for a mixed-objectives firm consists of the unique unit price $p_{nm}^{fb}(\theta)$ and transfer payment $T_{nm}^{fb}(\theta)$ satisfying

\[
p_{nm}^{fb}(\theta) = \frac{1}{1+\gamma} V_x(x^{fb}) - \frac{(1-\beta)}{\beta} \phi'(x^{fb}(\theta)),
\]

\[
T_{nm}^{fb}(\theta) = g(x^{fb}, \theta) - p_{nm}^{fb}(\theta)x^{fb},
\]

and for a pure output-maximizing firm

\[
\{p_{np}^{fb}(\theta), T_{np}^{fb}(\theta)\} \in \{\{p, T\} \mid 0 \leq p \leq g_x(x^{so}, \theta) \text{ and } T = g(x^{so}, \theta) - px^{so}\},
\]

for all $\theta \in \Theta$.

Proof. The payment policy for a mixed-objectives firm is derived by taking the FOC of the regulator’s problem and substituting in the firm’s first-order condition. The proof for the output-maximizing firm is in the appendix. \qed

Because price is not present in the demand function, the first-best solution simply equates the marginal benefit of additional consumption with the social marginal cost. More importantly, because the socially optimal level of output is determined only by the firm’s service quality, the first-best and socially optimal outcomes are equivalent and by simply modifying the way consumers pay for the service, the regulator may be able to improve the outcome.\textsuperscript{75} Notably, when the firm is a pure output-maximizer, the optimal contract is no longer unique, suggesting the regulator will have additional flexibility to induce its preferred outcome when the firm has superior information.

\textsuperscript{75}The observation that a nonmarketed good does not lead to an outcome distorted away from the social optimum was first made by Ma (1994) and Chalkley and Malcomson (1998b) using similar models.
2.4.2 Asymmetric Information about $\theta$

When information is asymmetric, the regulator’s problem is to maximize

$$
\max_{p(\theta), U(\theta)} \int_{\Theta} \left\{ V(x^*(p(\theta), \theta), \theta) - (1+\gamma)g(x^*(p(\theta), \theta), \theta) - \frac{(\lambda+\gamma)}{\beta} (U(\theta)-(1-\beta)\varphi(\theta)) \right\} dF(\theta),
$$

subject to individual rationality and incentive compatibility constraints similar to those for the marketed good.

The lack of a demand response to price also simplifies the regulator’s problem under asymmetric information because the SCP is automatically satisfied as reported by Lemma 5.

**Lemma 5.** When the good is nonmarketed, the SCP is satisfied and $d\{U_p/U_T\} / d\theta < 0$ for all $\theta \in \Theta$.

*Proof.* Taking derivatives of the firm’s profit with respect to price and transfer payment gives $U_p/U_T = x^*$. From the conjugate pairs theorem $\text{sign}[dx^*/d\theta] = \text{sign}[U_{x\theta}]$ and $\beta \Pi_{x\theta} = -\beta g_{x\theta} < 0$.

Any incentive compatible mechanism must still satisfy Lemma 4.$^{76}$ Because the MRS of the unit payment for the fixed transfer is increasing in the cost parameter, another simplification to the regulator’s problem due to the lack of price response is that condition (ii) of Lemma 4 reduces to the following.

**Lemma 6.** When the good is nonmarketed, the payment policy $\{p(\theta), T(\theta)\}_{\theta \in \Theta}$ is incentive compatible only if $dp/d\theta \leq 0$.

*Proof.* See Appendix B.  

$^{76}$Note that the proof for Lemma 4 holds when $\partial g/\partial p = 0$.  

87
Intuitively, the regulator must set a lower unit payment in higher cost states and compensate the firm via more of the fixed transfer to remove any incentive to misreport the cost as being higher than it is.

A Mixed-Objectives Firm

The regulator’s problem can again be solved utilizing optimal control. The first-order condition of the Hamiltonian yields

\[ V_x(x^*(p(\theta), \theta), \theta) - (1+\gamma)g_x(x^*(p(\theta), \theta), \theta) = (\lambda + \gamma) \left[ \frac{F(\theta)}{f(\theta)} g_{\theta x}(x^*; \theta) - \frac{1-\beta}{\beta} \phi'(x^*) \frac{dx^*}{dp} \right]. \]

(2.15)

The quantity \( x^* \) solving Eq. (2.15) is the second-best quantity given the regulator’s constraints. The interpretation of Eq. (2.15) is similar to the interpretation of Eq. (2.12), the first-order condition for a marketed good. When the firm has a mixed objective the direction of the distortion again depends on whether the firm’s information rents are increasing or decreasing with the unit payment and Proposition 9 continues to apply; however, the lack of a demand response to price simplifies the results in two ways. First, because \( dx^*/dp > 0 \) the output will always be distorted in the same direction as the unit payment as defined in Proposition 9. Second, because \( d\{g_{\theta}(x^*; \theta)\}/dp (= g_{\theta x}(dx^*/dp)) \) is unambiguously positive, the second-best unit price and equilibrium quantity are distorted strictly downward from the first-best levels with a pure profit-maximizing firm. This result is stated in the following proposition.
Proposition 14. When the good is nonmarketed and the firm is profit-maximizing, the second-best unit price and equilibrium quantity are distorted downward from the first-best for all but the lowest state, $\theta$ where they are equivalent.

Proof. The proof follows immediately from Proposition 9.

The removal of a demand response reduces the characteristics of the problem to rule (iii) of Proposition 10 and can be graphically represented by figure 2.1(c) where the level curves are transformed into vertical lines and the iso-rent curves continue to be increasing away from the origin.

Output-Maximizing Firm

When the firm’s preference for community benefit is so strong that it is a pure output-maximizer, then the payment rule established by Proposition 11 continues to be the unique payment rule inducing the first-best output when the good is nonmarketed. Moreover, as there is no demand response to the unit price, the equilibrium output is uniquely determined by quality. Thus, as long as $dp^{np}/d\theta < 0$ for incentive compatibility, the regulator is free to use the unit price to induce the firm to produce in the social interest and the payment rule reported by Proposition 11 induces the first-best outcome. This is stated in the following proposition.

Proposition 15. When the good is nonmarketed and the firm is output-maximizing the regulator can induce the first-best outcome using the payment rule reported by Proposition 11 with the exception that $p(\theta)$ may take a negative value.

Proof. See Appendix B.
2.5 Asymmetric Information about Demand

Lewis and Sappington (1992) find in a model without quality that asymmetric demand information does not result in an output distortion away from the first-best levels, in stark contrast to when asymmetric information is with respect to the firm’s cost. We investigate whether the present results are robust to the source of asymmetric information.

To see how asymmetric demand information alters the results we make the following changes to the model. First, the state parameter now represents a shock to demand instead of cost: \( x(q, p; \theta) \). To maintain consistency with the model with asymmetric knowledge of cost, we assume that higher states result in less demand at the same quality level and unit price: \( x_\theta < 0 \). As before we will work with the quality demand function \( q(x, p; \theta) \); therefore, because \( x_\theta < 0 \), it must be the case that \( q_\theta > 0 \) and in higher demand states the firm’s cost of production is higher for the same equilibrium quantity.

It will continue to be more convenient to reduce the problem by one dimension and work with quality-adjusted cost and social value functions. The quality-adjusted cost function is now defined as:

\[
g(x; p, \theta) = c(x, q(x, p; \theta)).
\]

Because higher demand-states soften demand, the partial derivatives and crosspartials are consistent across models: \( g_x > 0, g_p > 0, g_\theta > 0 \), and \( g_{\theta x} \geq 0 \).
The quality-adjusted social value function now takes the demand state as an argument,

\[ V(x; p, \theta) = B(x, q(x; p; \theta)) , \]

as it is now dependent upon the demand state. A higher state results in a left-ward shift in the demand curve so the social value to consuming \( x \) units must be lower in higher demand-states (i.e., \( V_{\theta} < 0 \)), and higher in low demand-states.

Condition \((i)\) of Lemma 4 identifies a necessary condition for an incentive compatible payment rule with asymmetric knowledge of cost. When the firm earns some information rent at an optimum (\( \Pi > 0 \)), the condition, \( dU/d\theta = -g_{\theta}(x; p, \theta) < 0 \), applies regardless of the source of asymmetric information. Similarly, condition \((ii)\) identifies a sufficient condition, which is dependent on the properties of the firm’s objective function. The properties of \( g \) do not change based on the source of information asymmetry, so this condition applies to both cases of asymmetric information. In consequence, the firm earns no rents in the highest state, \( \bar{\theta} \), and there is no output distortion in the lowest state, \( \theta \), regardless of the source of information asymmetry.

For a marketed good, the conditions that determine the relative size of the second-to first-best prices as enumerated by Proposition 10, are independent of the source of asymmetric information. Because the relationship between outputs is depends on how the unit payment affects the firm’s information rents, the source of asymmetric information can affect the direction of an output distortion. That is, the relationship between the second- and first-best outputs is dependent on the sign for \( d\{g_{\theta}\}/dp \),
which varies with the source of asymmetric information. This can be seen by decomposing $d\{g_\theta\}/dp$ into its component parts:

$$d\{g_\theta(x^*; p, \theta)\}/dp = g_{\theta p} + g_{\theta x}(dx^*/dp).$$

Regardless of the source of information asymmetry, $g_{\theta x} > 0$, and $dx^*/dp$ is unrestricted; however, $\text{sign}[g_{\theta p}]$ is somewhat dependent on the source of asymmetric information. When the asymmetry is in cost, $g_{\theta p} > 0$, but when it is in demand, the sign is ambiguous. This follows because with asymmetric knowledge of demand $g_{\theta p} = c_q q_\theta$ and $\text{sign}[q_{\theta p}]$ is unrestricted. When $q_{\theta p} > 0$, demand is more sensitive to the price in softer demand states, thus requiring ever increasing levels of quality to compensate for the demand response to an increase in price; and, when $q_{\theta p} < 0$ demand is less sensitive to price in softer demand states, reducing the rents the firm can extract, $g_{\theta p} < 0$.

When the asymmetry is with cost, then $g_{\theta p} > 0$ and the partial change in the firm’s rent with price and the partial change in the firm’s rent with output move in the same direction, i.e., $g_{\theta p}/g_{\theta x} > 0$. With asymmetric demand information the two may move counter to one another. When $g_{\theta p}/g_{\theta x} > 0$ then the relationship between first- and second-best prices and output are determined by Proposition 10. Proposition 16 identifies the relationship between first- and second-best prices when $g_{\theta p}/g_{\theta x} < 0$. 

92
Proposition 16. With asymmetric knowledge of demand, if \( \frac{g_{\theta p}}{g_{\theta x}} > 0 \), then the relative size of the second- to first-best price and output is identical to when the regulator’s uncertainty is in cost. Otherwise, the relative size of the second- to first-best price and output is determined by the rules:

\[
\begin{align*}
(i) & \quad 0 < -\frac{g_{\theta x}}{g_{\theta x}} < \frac{dx^*}{dp} \Rightarrow \frac{d}{dp} \left[ \frac{\partial U}{\partial \theta} \right] > 0 \Rightarrow \begin{cases} p^{sb}(\theta) < p^{fb}(\theta) \\ x^{sb}(\theta) < x^{fb}(\theta) \end{cases} \\
(ii) & \quad 0 < \frac{dx^*}{dp} < -\frac{g_{\theta p}}{g_{\theta x}} \Rightarrow \frac{d}{dp} \left[ \frac{\partial U}{\partial \theta} \right] < 0 \Rightarrow \begin{cases} p^{sb}(\theta) > p^{fb}(\theta) \\ x^{sb}(\theta) > x^{fb}(\theta) \end{cases} \\
(iii) & \quad -\frac{g_{\theta x}}{g_{\theta x}} < 0 < \frac{dx^*}{dp} \Rightarrow \frac{d}{dp} \left[ \frac{\partial U}{\partial \theta} \right] < 0 \Rightarrow \begin{cases} p^{sb}(\theta) > p^{fb}(\theta) \\ x^{sb}(\theta) < x^{fb}(\theta) \end{cases}
\end{align*}
\]

for all \( \theta \in [\bar{\theta}, \theta] \) and \( p^{sb}(\bar{\theta}) = p^{fb}(\bar{\theta}) \).

The intuition behind Proposition 16 follows similarly to Proposition 10. In case (i), \( dx^*/dp > -(g_{\theta p}/g_{\theta x}) \) and the second-best output must be under-supplied relative to first-best because the firm’s rents are increasing with the unit payment and the regulator will set a price below the first-best in order to limit the firm’s rents. Furthermore, because price and output are below first-best levels, quality must be unambiguously lower than the first-best level of quality. In case (ii), the firm still optimally increases output with a price increase, but the firm’s rents are decreasing faster with the change in output than they are increasing with a change in price resulting in a net decrease in rents with an increase in price. To limit the firm’s rents, the regulator will set a price above first-best, resulting in an oversupply. Moreover, because prices and output are above first-best levels, then it is clear that quality is unambiguously above the first-best level as well. Similarly, the firm’s information rent is decreasing in the price in case (iii), the difference is the firm optimally chooses a lower output quantity with a higher unit price resulting in an undersupply relative to first-best.
Lastly, because $g_\theta$ is independent of the source of information asymmetry, all of the results remain unchanged when the firm’s preference for community benefit is so strong that it is a pure output-maximizer. Reflecting this, the proofs include a parameterization that generically represents either asymmetric information for cost or demand.

2.6 Concluding Remarks

This chapter has examined the optimal payment policies for a monopolist who can manipulate demand through its choice of unverifiable quality. We have assumed that the regulator cannot contract on quality, output, or the firm’s cost ruling out many contracting regimes such as rate-of-return or minimum quality standards regulation. Moreover, to further complicate the regulator’s problem, we have assumed that the firm possesses superior information regarding some aspect of the market: either cost or demand. We have found that within the same informational environment the regulator can achieve strikingly different outcomes based on the consumers’ access to the good and the firm’s objective. Indeed, when the good is nonmarketed and the firm is a pure output-maximizer the regulator can completely attenuate the informational advantage of the firm. Because there is a deadweight loss associated with using public funds, however, eliminating the firm’s informational advantage includes a social cost and does not represent a panacea for the regulator. In contrast, when the good is marketed the equilibrium output may be under- or oversupplied relative to the social optimum — even for a pure profit-maximizing firm — in strong contrast to the previous regulatory literature.
The ambiguous direction of the distortion that occurs with a marketed good follows directly from the fact that the firm can manipulate demand through its choice of quality. If the regulator raises the unit payment, then the firm adjusts quality to compensate for the negative demand response, re-optimizing its choice of output. The regulator wants to use the payment policy to discipline the firm to truthfully reveal its private information, and not extract any more of an information rent than necessary. To this end, the regulator may need to either discipline the firm by setting the price above the first-best level to make it more expensive to produce output, or by setting the price below first-best to decrease the firm’s unit revenue. In contrast, when the firm cannot manipulate demand by adjusting quality (because it is either not in the model, or consumer demand is inelastic to quality), then the unit price uniquely determines the quantity demanded. The distortion from the social optimum caused by the firm’s informational advantage is always downward in this case as the regulator must shade the unit payment in order to extract some of the firm’s information rent. This result is also analogous to that found in single unit procurement models because demand is price-inelastic and higher payments always increase the firm’s profit.\footnote{For example Baron and Besanko (1984) and Laffont and Tirole (1986).}

The results of the model represent only the best possible outcomes. Given the complex interactions of the cost, value, and demand functions, for many classes of functions, the optimal payment policies for the various scenarios considered may not satisfy incentive compatibility. When they do not, the regulator will not be able to extract any information from the firm and is better off setting a constant payment.
rule, resulting in strong distortions away from the social optimum regardless of the type of good and objective of the firm. The optimal payment policy need not exhibit complete separation or complete pooling either, but instead there may be pooling for some subset of cost or demand states and separation for others and only when the functional forms are known can we identify if the regulator can achieve the second-best outcome.

The results of this chapter can help guide policy makers to determine to what degree the non-contractibility of quality is a problem. For example health policy makers have recently taken a strong interest in reorganizing physicians and hospitals into Accountable Care Organizations (ACOs). The objective is to promote the socially preferred level of quality of care while simultaneously introducing incentives to lower costs. High-powered contracts that pay a fixed capitation per patient provide a strong incentive to reduce costs, but it is not clear how quality may be affected and to what degree the government should expend resources to quantify quality. This research provides testable predictions showing that the answer depends on the relative responsiveness of consumers to quality and the providers’ cost of producing quality. Moreover, it provides guidance into the value of increasing consumer responsiveness to quality through such programs as hospital and HMO report cards.

Before concluding, we highlight two important directions for future research. First, in analyzing the effect of the consumers’ incentive response to the contracted unit price we took the extreme position that either the regulator or the consumers are responsible for the entire payment; however, in many regulated markets the government and
consumers share responsibility. For example, in voucher programs consumers are provided a voucher for tuition at the school of their choice but schools are not limited to charging the voucher amount and consumers may have to kick in a payment above the voucher. In this way the voucher softens the consumers’ price elasticity of demand, but does not make it completely inelastic. Similarly, as a part of the PPACA the government has mandated insurance coverage. To help those for whom premiums would exceed a certain percentage of income, the government provides subsidies, softening the price elasticity of demand for those eligible. Given the prevalence of such mixed payment systems, studying the optimal tiered payment policy is an interesting and important avenue of future research.

Secondly, despite assuming a market environment in which the regulator cannot observe the firm’s costs, the output, or quality level, the information burden on the regulator is still exceedingly high. The regulator is assumed to know or have a strong prior for the firm’s cost of production and the characteristics of consumer demand. Since regulators are generally much less informed about the details of the firm’s technology, especially in an environment such as health care where those technologies are evolving rapidly, it will be beneficial for future research to consider the nature of regulatory policies under even more restricted information regimes.
Chapter 3: Diagnosing Hospital System Bargaining Power in Managed Care Networks

There has long been a concern that hospital system growth through mergers and acquisitions has generated increased market power for system hospitals resulting in higher reimbursement rates. Economists and antitrust authorities traditionally approached hospital systems as classical horizontal mergers, analyzing the relationship between hospital market concentration and prices. More recently, with the rise of managed care, the literature has shifted its focus toward the bilateral bargaining game between hospitals and managed care organizations (MCOs). Since MCOs are the primary purchasers of hospital services, a hospital’s market power is largely determined by how much the MCO’s enrollees value the inclusion of the hospital in their network. Hospitals that have fewer local competitors or that provide a unique set of services generate a larger value (i.e., incremental surplus) to the MCO and can command higher reimbursement rates because the services of alternative hospitals are not viewed as a close substitute. Consequently, if two local hospitals merge to form a system and negotiate as one with MCOs, the incremental value of including the two-hospital system in the network is now larger because the MCO can no longer use one hospital as a substitute for the other. In other words, the increase in market power is
a direct result of an increase in the size of the surplus over which the MCO and the two hospitals bargain.\textsuperscript{78} Much of the recent empirical work on hospital competition focuses on identifying this local market power effect.\textsuperscript{79}

Our study investigates an additional, important channel through which system formation may impact reimbursement rates: by altering the relative bargaining power of hospitals vis-a-vis MCOs. Local market power allows system hospitals to bargain with MCOs over a larger surplus because their incremental value to the MCO network is larger. It is also possible that system hospitals are able to extract a larger share of this surplus from MCOs by leveraging the system in some way to negotiate more favorable rates.

One possibility is that hospitals in systems are able to share the costs of creating a larger and more skilled team of contract negotiators. For example, in 2004 Tenet Healthcare—a national system of 73 hospitals—adopted a “national negotiating template and new technology to analyze payer-specific profit and loss data, giving negotiators ammunition during contract talks (Colias, 2006).\textsuperscript{80} Individual hospitals may not have the size or resources to pursue such strategies.

Alternatively, seminal works by Binmore, Rubinstein, and Wolinsky (1986) and Rubinstein, Safra, and Thomson (1992) show that in a two-player bargaining game

\textsuperscript{78}The surplus created by including the system in the MCO network is larger than the surplus created by adding both hospitals individually to the network prior to the formation of the hospital system because consumers can experience a significant reduction in utility when both hospitals are removed from their network even if removing either hospital individually has very little impact on utility.

\textsuperscript{79}For example, see Town and Vistnes (2001), Capps et al. (2003), and Ho (2009).

\textsuperscript{80}According to Tenets CEO this initiative was necessary because Tenet was “being outgunned by the managed care companies in negotiations(Colias, 2006).
the more risk averse player receives a lower share of the net surplus (i.e., higher relative risk aversion lowers bargaining power). Negotiations with system hospitals can pose greater risk to MCOs if a breakdown in bargaining leads all of the systems hospitals to withdraw from the MCO’s network. When MCO executives or their contract negotiators are risk averse (or are more risk averse than hospital system executives), they may be at a greater disadvantage in negotiations with system hospitals than with non-system hospitals.

In these cases, mergers of hospitals within the same patient market can result in significant changes to both the bargaining power and local market power of the hospitals. Moreover, the formation of hospitals systems across geographic markets that have no impact on local market concentration could still have significant effects on bargaining power and result in higher reimbursement rates. Such bargaining power effects have been overlooked by existing analyses that focus only on local market concentration.

To better identify the differences in reimbursement rates that arise because of the market power effect from those that arise because of variation in the bargaining power of system and non-system hospitals we utilize an empirical strategy that identifies bargaining power as a function of hospital and system characteristics. This is accomplished by first estimating a structural model of demand based on the observed characteristics of patients, their illnesses, and of hospitals. The incremental value of adding a hospital to an MCO’s network is derived from the estimated utility of all patients who would choose the hospital after having become ill (Capps, Dranove, and Satterthwaite, 2003). Additionally, the demand estimates are used to both predict
the number of additional managed care patients that will visit a hospital if it is added to an MCO’s network as well as to predict where patients would alternatively go if a given hospital is not in their choice set.

Next, using data on hospitals’ operating costs, in-patient days, out-patient visits, and other characteristics we estimate a hospital cost function. The cost function is used to estimate the change in cost for a hospital that follows from treating the MCO’s patients that are predicted to visit the hospital. The producer surplus that is shared between a hospital and MCO is calculated using the incremental value of adding a hospital to an MCO’s network net the hospital’s expected cost of treating the patients enrolled with the MCO as well as the change in the MCO’s reimbursement expenditures that are a consequence of patients choosing the hospital over alternative hospitals within the network. Finally, we incorporate data on hospital revenues into an asymmetric Nash bargaining model to estimate how the available surplus is split between hospitals and MCOs. Hospital bargaining power is specified as a function of various hospital, system, and market characteristics to identify the factors associated with differences in bargaining power.

Our main results indicate that systems that alter the concentration of hospitals in a local patient market extract a significantly higher proportion of the available surplus; i.e., have significantly higher bargaining power. For instance, hospitals having a close (≤ 2.5miles) system partner have substantially more bargaining power than a hospital without a close partner. Additionally, a hospital’s bargaining power is increasing with the market share of its system within the local patient market around that hospital, even after controlling for the market-power of the system. These findings
are a clear indication that concentration within a local patient market is one of the most important determinants of a hospital’s bargaining power vis-á-vis an MCO as it makes that system and its hospitals more valuable, or indispensable, to an MCO. Conversely, a higher concentration of MCOs within a hospital’s patient market lowers the bargaining power of the hospital.

We try several different specifications to determine if a system can use its size in some way to increase the bargaining power of a member hospital. For example we find some evidence that hospitals belonging to systems operating in multiple patient markets are associated with higher bargaining power—particularly for-profit hospitals. However, we find no evidence that system size—measured by the number of beds or number of systems—is correlated with bargaining power. Many system characteristics such as system market share, the presence of a close partner, and belonging to a multi-market system are all highly correlated with system size so are likely picking up any differences in bargaining power associated with larger systems.

In addition to system membership we find that other hospital characteristics are correlated with higher bargaining power. For example rural hospitals have less bargaining power than urban hospitals and, lending support to the predictions of Gal-Or (1999), hospitals with physician arrangements such as a Group Practice Without Walls (GPWW) also have significantly more bargaining power than hospitals without such arrangements.81 This latter finding is particularly troubling given the move by policy

---

81 A GPWW is a legal entity formed by a network of physicians who maintain their individual practice locations but share administrative or ancillary expenses. Often times a GPWW is sponsored by a hospital suggesting a tighter level of integration of coordination between the GPWW and hospital (American Hospital Association).
makers to move toward more vertical integration between physicians and hospitals through the formation of Accountable Care Organizations.

The paper develops as follows. We begin by providing some background on the evolution of the hospital-MCO relationship in Section 3.1. Section 3.2 provides a summary of the related literature, highlighting the differences in the current study and previous studies. Section 3.3 develops the estimation strategy. Section 3.4 describes the data. Results to the demand, cost, and bargaining estimations are presented in Section 3.5. In Section 3.6 we decompose the difference in reimbursement rates between system and non-system hospitals into the portion that is attributable to differences in market power and the portion that is attributable to differences in bargaining power. Lastly, Section 3.7 ends with some concluding remarks.

3.1 Background

The relationship between hospitals and insurers has undergone significant changes in the last two decades. Before MCOs—such as health maintenance organizations (HMOs)—came to prominence hospitals were paid on a fee-for-service basis yielding substantial power to hospitals to set their own prices. With the introduction of HMOs, however, the nature of hospital pricing and competition has changed from being patient-driven to being payer-driven (Dranove et al., 1993).

3.1.1 Market Structure

In the early 1980s and 1990s many HMOs contracted with less than half of the hospitals in their markets (Feldman et al., 1990) and this selective contracting gave
them the ability to steer patients away from one hospital to another. HMO penetration rates steadily increased through the 1990s, increasing from 16 percent in 1988 to 31 percent in 1996 when they reached their peak. As a consequence of the selectivity and market share growth of HMOs, hospitals were forced to become more competitive with their prices in order to secure HMO membership or risk substantial reductions in patient volume.

Managed care has continued to evolve since the introduction of HMOs and now includes Preferred Provider Organizations (PPOs) and Point of Service (POS) plans. Like an HMO, PPOs negotiate reimbursement rates with hospitals; however, unlike an HMO, an enrollee to a PPO is free to visit any hospital, in- or out-of-network. The enrollee is discouraged from visiting out-of-network hospitals, however, by having higher out-of-pocket costs when he does, giving a PPO some ability to steer patients to in-network hospitals. POSs are less common and considered a hybrid between an HMO and a PPO.

Consumers have come to prefer PPOs over HMOs and they are now the predominant form of not just managed care, but of insurance in general. As an example of the sensational growth of PPOs, in 1999 PPOs accounted for 39% of the employer sponsored insurance market whereas HMOs accounted for 28%, POSs accounted for 23%, and traditional indemnity plans accounted for 10%. Just ten years later, though, PPOs had grown to account for 60% of the market, with HMOs accounting for 20%, POSs accounting for 19%, and indemnity falling to an almost nonexistent 1% of the employer-based insurance market (Claxton et al., 2009). Despite whether an MCO takes the form of a HMO, PPO, or POS, hospitals have an incentive to join the MCO’s
network and the fact that managed care now represents the overwhelming majority form of insurance results in an enormous shift in bargaining power towards insurers relative to the pre-managed care era.

In addition to hospitals losing bargaining power as insurers switch to managed care, the insurance market has seen a rapid amount of consolidation further eroding hospital bargaining power. For example, according to the American Medical Association’s Division of Economic and Health Policy Research, in 2006 18 states had two insurers controlling 70 percent of the managed care market, and a year later the number of states with such high levels of insurer concentration increased to 24.

As a potential response to the increasing market power of insurers, hospitals have also undergone a tremendous number of mergers and acquisitions over the last two decades. For example in the 10 year period of 1998 to 2008, there were an average of 73 hospital mergers and acquisitions every year.\footnote{American Hospital Association Trendwatch Chartbook, 2009; http://www.aha.org/aha/research-and-trends/chartbook/2009chartbook.html} The result has been an increasing prominence of hospital systems in which hospitals are either jointly owned or contract managed by a third-party. Today over 57% of all acute-care hospitals in the U.S. are in a system.\footnote{See Cuellar and Gertler (2003) for a more detailed look at the growth and breakdown of system membership.}

The tremendous amount of consolidation in the acute-care hospital market raises several potential concerns. First, increased market share in a specific patient market may generate higher local market power. In response to this concern, antitrust
authorities have traditionally focused mostly on hospital concentration within a local patient market when deciding whether or not to challenge the merger. However, systems operating in multiple patient markets may also be able to leverage their size across patient markets to negotiate higher prices, as Dranove and White (1998) first point out, this suggests the need to move beyond the traditional definition for a hospital’s market in assessing system formation and growth.

### 3.1.2 Contract Negotiations

Whether hospitals and MCOs are negotiating a contract for the first time or renegotiating a contract, the negotiation process can be very complicated. Hospitals and MCOs must determine the expected case-mix across clinical areas, the treatment settings, quality targets, and historical payer behavior with respect to claims denial and underpayments (Boyd and Finman, 2010; Osten, 2011). Furthermore, payments often take many different forms ranging from capitation, to per diems, to fee-for-services reflecting the relative risks and preferences for taking on those risks. Contracts may specify distinct payment rates for thousands of different Current Procedural Terminology (CPT) codes and can be hundreds of pages in length. The complexities of the contract require negotiators to have a substantial amount of information and skill in order to achieve a favorable outcome. In fact Kongstedt (2001) claims that the skills of the negotiator are the most important component in the negotiating process.

Given the complexities of the contract and the importance of the negotiators’ skills, it would not be surprising to find significant differences in relative bargaining
power across hospitals. Articles in the relevant trade media often highlight specific factors that can influence a hospital’s bargaining power over an MCO. For example, it is suggested that the importance of a hospital to an MCO (as measured by the share of that MCO’s enrollees that will prefer that hospital) will give it power to negotiate higher rates (D’Cruz and Welter, 2008; Lowes, 2008). Hospitals that provide more effective care may also be able to secure higher rates beyond the increase they could receive due to higher demand from patients because they may lower the overall costs of treatment for the MCO through reduced complications and utilization (Lowes, 2008). Lastly, the ability of a hospital to credibly threaten to walk away from the negotiation can leave it in a stronger bargaining position (Lowes, 2008; Rollins, 2000). These three factors in particular are traits that system hospitals could have in common through their ability to utilize or leverage the entire system.

3.2 Related Literature


Grennan (2010) identifies similar differences in hospitals bargaining power in the negotiations with suppliers for the purchase of coronary stents.
a system control in addition to system concentration indices. These analyses consist of regressing some measure of a hospital’s price against a measure of market concentration—generally a Herfindahl-Hirschmann index (HHI)—and hospital characteristics. A shortcoming of these reduced form studies is the inherent difficulty with disentangling a price differential from unobservable differences in quality and/or costs from that of market-power. If a hospital’s market share is based on a higher service quality, then its price differential may be incorrectly attributed to market power. Importantly, however, Melnick and Keeler (2007) find that even after controlling for the increased market concentration created by systems, hospitals belonging to system hospitals operating in multiple patient markets have even higher prices.

Brooks, Dor, and Wong (1997) are the first to apply a structural bargaining model to the hospital pricing problem. Brooks et al. utilize an asymmetric Nash bargaining model and focus on the negotiated prices for an appendectomy. The authors do not control for quality differences between hospitals and the surplus, from which the hospital’s profit is extracted, is defined as the difference in the hospital’s list prices and the Medicare reimbursement rate. This can be problematic, though, if list prices have little correlation with the value of the hospital to patients and the MCO. Moreover, the bargaining model utilized assumes that there is only one hospital and one insurer in the market generating a different surplus calculation than if there are multiple hospitals.85 Similarly Halbersma, Mikkers, Motchenkova, and Seinen (2010) utilize the bargaining model of Brooks et al. (1997) and examine the bargaining power of hospitals in the Netherlands.

85 As a consequence to there being only one hospital, if the hospital and MCO fail to negotiate a contract, patients will still visit the hospital and the MCO must pay the hospital’s list prices.
Town and Vistnes (2001) take the next step in analyzing hospital-MCO bargaining by modeling the hospital’s reimbursement rate as a function of the incremental value of a hospital to the MCO as well as the incremental value of the next closest substitute hospital capturing the effect of strategic HMO network formation has on a hospital’s price. Town and Vistnes use data from two HMOs operating in California to estimate how various hospital characteristics affect the per diem price it receives. System status is not considered, although a simulation is performed to quantify the degree to which prices change when hospitals, which are necessarily in the same market, merge.

Capps et al. (2003) build off of Town and Vistnes (2001) by using the option value estimates to calculate the average bargaining power of hospitals using data for the San Diego market. Although we utilize the option-demand framework developed by Capps et al., there are a couple notable methodological differences between the two analyses. First, Capps et al. assume that there is no change in cost for the MCO when a hospital is added to its network. We identify the MCO’s change in cost by using the demand estimates to simulate the change in the MCO’s expenditures resulting from adding a hospital causing some of its enrollees to switch from other in-network hospitals to the added hospital. Second, Capps et al. assume that there is no change in cost to the hospital from joining a network. These assumptions do not allow for separate identification of differences in bargaining power, which are important contributors to the price differential between system and non-system hospitals.

Both assumptions essentially follow the bargaining model of Brooks et al. (1997) and are based off of the assumption that enrollees still visit the hospital when the hospital and MCO fail to negotiate a contract.
Our model is most closely related to Grennan (2010) who utilizes a Nash bargaining framework to analyze the welfare effects of price discrimination of medical device manufacturers in the market for coronary stents. Grennan also separately identifies the market power and bargaining power of hospitals with respect to the medical device makers. Moreover, Grennan allows bargaining power to vary by hospital reflecting idiosyncratic difference in the hospitals’ negotiating skills. However, as the focus of the paper is on the welfare effects of price discrimination, Grennan does not identify hospital or system characteristics that may be associated with bargaining power whereas we allow bargaining power to vary based on observable characteristics of the hospital and its patient market allowing us to decompose the price differences between system and non-system hospitals into differences in market and bargaining power.

Gaynor and Vogt (2003) take a different structural approach by modeling hospital markets as a differentiated product oligopoly and estimate the market supply and demand for hospital services in order to simulate hospital system mergers. One focus of Gaynor and Vogt (2003) is on the difference between for-profit and not-for-profit hospitals in which they conclude that there is no difference in their willingness to exploit market power. This approach assumes a traditional product market where firms compete for customers via price and quality competition. However, because of insurance, and in particular because of managed care, we believe that hospitals compete for patients through quality competition, but prices are determined through a bilateral bargaining game and not through a Bertrand pricing game. Moreover, the
differentiated product oligopoly model cannot identify the rents hospital systems may be able to extract when they operate in multiple patient markets.

Utilizing a moment inequality approach, Ho (2009) uses data on MCO networks and premiums to model the hospital-MCO bargaining process. She finds that system hospitals, capacity constrained hospitals, and star hospitals\textsuperscript{87} attain higher mark-ups over cost. Although markups are related to bargaining power, they are not identical. For example, if two hospitals have identical costs, but one is more attractive to consumers, allowing the MCO to charge higher premiums, then the process of dividing that surplus between the MCO and hospital can result in higher markups over cost with the same division of surplus. Moreover, higher markups can occur through the market-power effect that increases the value of system hospitals to MCOs with no change in bargaining power, thus making the two effects indistinguishable. Ho also takes a different approach to analyzing the bargaining problem as a game of simultaneous take-it-or-leave-it offers from hospitals. Identification comes from the variation in probability of agreement across different types of hospitals and plans, which necessarily limits the analysis to HMOs and POSs with restrictive networks since PPOs generally contract with all, or nearly all, of the providers in a market. Furthermore, the approach is computationally intensive, limiting the number of covariates that can be identified. This limitation does not allow for identification of the characteristics of a hospital system that are associated with bargaining power.

In addition to examining what hospital characteristics lead to higher reimbursement rates, a couple studies have examined what MCO characteristics lead to lower

\textsuperscript{87}Star hospitals are defined as those hospitals which have market shares above the ninetieth percentile when MCOs contract with all hospitals in the market.
reimbursement rates. For example Sorensen (2003) examines how an HMO’s ability to channel patients affects its ability to secure lower reimbursements. Using the HHI for MCOs, Sorensen finds that a one standard deviation increase in HHI leads to a 13 percentage point drop in the reimbursement rate. In a similar study Wu (2009) considers the degree to which the hospitals’ cost differential, the ability of MCOs to channel patients, and the excess capacity of hospitals contributes to the per diem price differential of hospitals. A hospital’s ability to channel patients is measured by first estimating a hospital choice model using a conditional logit and using the difference between a patient’s predicted choice and his actual choice. Wu finds that channeling and size both allow an MCO to secure larger discounts, but the magnitudes are quite small. For example, she finds that a one standard deviation increase in payer size leads to a 1 percentage-point decrease in reimbursement; and a one standard deviation increase in ability to channel patients results in a 2 percentage point drop in reimbursements.

3.3 Model for Estimation

3.3.1 Willingness-to-Pay

We utilize the framework developed by Capps et al. (2003) to estimate the market willingness-to-pay for having access to a hospital.\footnote{Town and Vistnes (2001) and Ho (2009) calculate the value of a hospital to an MCO’s network in a similar manner.} We begin by modeling the utility for patients receiving treatment at a hospital. The \textit{ex post} utility of patient $i$ who,
after developing an illness, obtains treatment from hospital \( h \) is defined as

\[
U_{i,h}(H_h, X_i, \lambda_i) = \alpha R_h + H_h' \Gamma X_i + \beta_1 T_h(\lambda_i) + \beta_2 T_h(\lambda_i) \cdot X_i + \beta_3 T_h(\lambda_i) \cdot R_h - \gamma(X_i) P_h(Z_i) + \epsilon_{ih},
\]

where \( H_h = [R_h, S_h] \) is a column vector of hospital \( h \)'s characteristics which are common across all illnesses, \( R_h \), and those characteristics which are illness specific, \( S_h \). \( X_i = [Y_i, Z_i] \) is a column vector of the patient’s characteristics, \( Y_i \), such as age, race, and gender as well as clinical attributes, \( Z_i \), such as diagnostic category, the presence of complications and whether or not the treatment is surgical. \( T_h(\lambda_i) \) is the approximate travel time from patient \( i \)'s location \( \lambda_i \) to hospital \( h \). The function \( \gamma(X_i) \) converts money into *utils* for a patient with characteristics \( X_i \) and \( P_h(Z_i) \) are the out-of-pocket costs for patient \( i \) having clinical attributes \( Z_i \) at hospital \( h \). Lastly, the error term \( \epsilon_{ih} \) is assumed to be an i.i.d. extreme value random variable representing the idiosyncratic component to patient \( i \)'s utility for being treated at hospital \( h \).\(^{89}\)

Hospital characteristics, which may affect the quality of its services, include properties such as teaching status, for-profit status, system and network membership and ownership by a physician group. The quality of care delivered by a hospital may also be affected by any physician arrangements used as part of an integrated healthcare delivery program, especially when those arrangements are selective about which physicians may become members; therefore we include indicators for four common

\(^{89}\)Due to dimensionality issues we do not include a full set of hospital effects. We control for several observable quality characteristics as well as the systematic, unobservable differences that are attributable to for-profit status, teaching status, rural status, and system membership, thus the unobservable quality differences, \( \nu_h \), are likely to be small if they are captured by these other observable characteristics. See Tay (2003) for a more thorough discussion of the issues of hospital quality differences as well as issues of endogeneity. In a similar model for hospital demand, Ho (2006) finds that assuming \( \nu_h = 0 \) does not lead to significant bias in the demand estimates.
physician arrangements. A hospital’s services include items such as high-technology imaging equipment and service-specific items such as a birthing room or the ability to perform heart surgery.

Given the ex post utility of patient $i$, as established in the extant literature on choice models, patient $i$’s interim utility of having hospital $h$ in his choice set $\mathcal{M} = \{1, 2, \ldots, M\}$ is denoted as

$$V(\mathcal{M} | H_h, X_i, \lambda_i) = E_{h \in \mathcal{M}} \max \left[ U(H_h, X_i, \lambda_i) + \epsilon_i \right] = \ln \left[ \sum_{h \in \mathcal{M}} \exp \{U(H_h, X_i, \lambda_i)\} \right].$$

(3.1)

From (3.1) it is clear that hospital $h$’s contribution to patient $i$’s interim utility derived from MCO $m$’s network $\mathcal{M}$ is

$$\Delta V_h(\mathcal{M} | H_h, X_i, \lambda_i) = V(\mathcal{M} | H_h, X_i, \lambda_i) - V(\mathcal{M} \setminus h | H_h, X_i, \lambda_i)$$

$$= \ln \left( \frac{1}{1 - s_h(\mathcal{M} | H_h, X_i, \lambda_i)} \right).$$

(3.2)

where $s_h(\mathcal{M} | H_h, X_i, \lambda_i)$ is hospital $h$’s market share when included in network $\mathcal{M}$ given by the logit demand specification:

$$s_h(\mathcal{M} | H_h, X_i, \lambda_i) = \frac{\exp \{U(H_h, X_i, \lambda_i)\}}{\sum_{j \in \mathcal{M}} \exp \{U(H_j, X_i, \lambda_i)\}}.$$

There is no outside option because the data contain only those patients which have become sufficiently ill that they choose to visit a hospital. Integrating (3.2) over the population distribution of patient attributes, diseases, and patient locations produces the ex ante value of including hospital $h$ in the network $\mathcal{M}$. Let $F(X_i, \lambda_i)$ denote the joint cumulative distribution of patient characteristics, diseases, and locations of all

90Independent Practice Association (IPA), Group Practice without walls (GPWW), Open physician-hospital organization (OPHO), and Closed physician-hospital organization (CPHO).
patients who will visit a hospital, then the total \textit{ex ante} WTP for inclusion of hospital $h$ in network $\mathcal{M}$ is

$$\Delta W_h(\mathcal{M}) = N \int_{X,\lambda} \frac{1}{\gamma_p} \ln \left( \frac{1}{1 - s_h(\mathcal{M} | H_h, X_i, \lambda_i)} \right) dF(X_i, \lambda_i),$$  \hspace{1cm} (3.3)

where $N$ is the number of consumers sufficiently ill that they visit a hospital in the choice set and $\gamma_p$ is the (assumed) constant conversion factor for converting dollars into \textit{utils}.\textsuperscript{91}

### 3.3.2 Hospital Cost

Following earlier literature on estimating the cost function of multi-product firms (See, for example, Fournier and Mitchell (1997), Bamezai and Melnick (2006), and Capps, Dranove, and Lindrooth (2010) for examples of applying the trans-log specification to hospital cost estimation) we use a form of the trans-log specification where hospital $h$’s cost at time $t$ is estimated as

$$\ln(Cost_{ht}) = \alpha_0 + \beta_Y \ln(Y_{ht}) + \beta_{YY} \ln(Y_{ht}) \times \ln(Y_{ht})$$

$$+ \beta_{WW} \ln(W_{ht}) \times \ln(W_{ht}) + \beta_{YW} \ln(Y_{ht}) \times \ln(W_{ht}) + H_h + \mu_t + \epsilon_{ht}.$$  \hspace{1cm} (3.4)

In Eq. (3.4) $Y$ is a vector of the hospital’s $m$ outputs, $W$ are the hospital’s $n$ inputs, $H$ is a vector of hospital characteristics, $\mu_t$ are time fixed effects, and $\epsilon_{ht}$ is a mean zero error term. A hospital’s outputs are defined as its inpatient days divided into Medicare, privately insured, and all other payers (Medi-Cal, worker’s compensation, etc), as well as outpatient visits also divided into Medicare, private managed care, private indemnity, and other payers. We use a day of inpatient care instead of a

\textsuperscript{91}Capps et al. (2003) provide a detailed discussion of how a variable $\gamma_p$ may bias the estimates.
discharge as a measure of output because the number of days of care for a given discharge correlates with the total cost of care more than any other available measure of severity.\textsuperscript{92} This is particularly important given that we do not have sufficient financial data to break down outputs by DRG, which would better control for the differences in costs across diagnosis. Inpatient days for different payers are treated as separate outputs to account for any case-mix and treatment intensity differences that may affect a hospital’s cost of treatment based on the patient’s insurer.

A hospital’s staffed beds multiplied by the nurse-to-bed ratio are used as a single input and captures both the size of the hospital and the efficiency of the hospital staffing. To some degree this also controls for case-severity mix as hospitals that generally admit sicker patients may have higher relative staff levels in order to meet the patients’ needs. We control for wage differences between hospitals by including fixed effects for the hospital’s health service area (HSA) as defined by the Centers for Disease Control. In order to measure short-run costs we do not include hospital fixed-effects (Baltagi and Griffin, 1984). However, to minimize omitted variable bias we include other hospital characteristics such as the hospital’s type of control (for-profit, non-profit), rural status, teaching status, if the hospital is operated by a physicians’ group, the ratio of Medicare to other payer discharges, the presence of a positron emission tomography machine, and the presence of a magnetic resonance imaging (MRI) machine.

\textsuperscript{92}We examined using the number of diagnosis or the presence of complications to produce a severity index, but length of care was found to be the best predictor of cost.
3.3.3 Bargaining

Similar to Brooks et al. (1997) and Grennan (2010) we utilize the cooperative bargaining model of Nash (1950, 1953) with the inclusion of an exogenous bargaining power parameter first introduced by Svejnar (1986). The bargaining problem for MCO \( m \) and hospital \( h \) can be expressed as

\[
\max_{p_{hm}} \left[ \Pi_m(M) - \Pi_m(M \setminus h) \right]^{1-\alpha_h} \left[ \Pi_h(H) - \Pi_h(H \setminus m) \right]^{\alpha_h},
\]

(3.5)

where \( \alpha_h \) is hospital \( h \)'s bargaining power vis-à-vis MCO \( m \) and \( p_{hm} \) is the reimbursement price agreed to by MCO \( m \) and hospital \( h \). \( \Pi_m(M) \) is MCO \( m \)'s profit from network \( M \) and the vector of reimbursement prices \( P_M = \{p_1m, p_2m, \ldots, p_{hm}, \ldots, p_{Mm}\} \) and \( \Pi_m(M \setminus h) \) is MCO \( m \)'s profit when it does not include hospital \( h \). \( \Pi_h(H) \) is hospital \( h \)'s profit when it has contracted the reimbursement price vector \( P_h = \{p'_1h, p'_2h, \ldots, p_{hm}, \ldots, p'_{Hh}\} \) from the set of \( H \) networks, and \( \Pi_h(H \setminus m) \) is hospital \( h \)'s profit when it is not in MCO \( m \)'s network.

Thinking of the bargaining problem as a strategic game of counter-offers, the disagreement point is defined as the payoff each party receives when they fail to reach an agreement (Binmore et al., 1986). For each hospital \( h \) negotiating with MCO \( m \) the disagreement point is \( \Pi_h(H \setminus m) \) as it is assumed that the hospital successfully negotiates contracts with the other MCOs operating in the patient market; and the disagreement point for MCO \( m \) is \( \Pi_m(M \setminus h) \) as the MCO is also assumed to successfully negotiate contracts with all other hospitals.\(^{93}\) If, however, a hospital is in a system and threatens removal of all system hospitals from MCO \( m \)'s network when

\(^{93}\)We are thus assuming the contract equilibrium introduced by Cremer and Riordan (1987).
a favorable reimbursement rate is not agreed upon, then the appropriate disagreement point for the MCO is \( \Pi_m(\mathcal{M} \setminus S_h) \) where \( S_h \) is the set of hospitals belonging to the same system as hospital \( h \). We provide estimates of hospitals’ bargaining power when they can threaten their removal from a network only and when they can threaten removal of the entire system.\(^{94}\)

The assumption that an MCO adds every hospital to its network, and every hospital is in each MCO’s network is especially appropriate for PPO networks. For instance a former contract negotiator with a major national MCO told us that because of competitive pressures their objective was to get nearly every hospital in their PPO network, and that he believed that was the other MCO’s objectives as well. Given these objectives, the bilateral bargaining game is a good way to model the hospital-PPO contract negotiation. This argument is slightly more problematic for HMO and POS networks since there may be some strategy involved with excluding hospitals. Even among these restrictive networks, however, Ho (2006) observes that on average 87% of hospital-MCO pairs establish a contract when the MCOs are HMOs or POSs.\(^{95}\)

Because the bargaining outcomes are private information, we assume that the bargaining outcome between hospital \( h \) and MCO \( m \) does not influence the bargaining outcome between hospital \( h \) and any other MCO \( m' \), or any other hospital \( h' \) and

\(^{94}\)In the former case system membership generates no market power as there is no difference in the incremental value of adding the hospital to a network and there is no change in the disagreement point from belonging to a system placing all of the difference in profit as a change in bargaining power.

\(^{95}\)Alternatively, our results could be interpreted as describing the differences in hospital bargaining power vis-à-vis PPOs as long as the overall average MCO reimbursement rate that we use is a close approximation of the average PPO reimbursement rate.
MCO $m$.\textsuperscript{96,97} In defining the disagreement points we must make an assumption regarding the choice patterns of consumers. For example, some consumers may have a strong preference for a hospital such that if their MCO drops that preferred hospital from its network then the consumer will switch MCOs. In contrast others may have a strong preference for an MCO and remain with their MCO if a preferred hospital is removed from its network. Because a majority of private insurance is provided by employers, limiting the ability of consumers to switch MCOs, we assume that in the short-run consumers are tied to the MCO. This assumption affects the disagreement points for the hospital and MCO as follows. If a hospital does not contract with a particular MCO then it is assumed to not receive any patients from that MCO lowering the demand for that hospital; and, if a MCO does not contract with a particular hospital, then the enrollees that would visit that hospital upon falling ill will instead visit a different hospital within the network.\textsuperscript{98}

\textsuperscript{96}The reimbursement rates established between an MCO and hospital are private information. MCOs and hospitals, however, can probably observe which other MCOs and hospitals have successfully negotiated contracts. This information may play a strategic role in the bargaining game, however since we do not observe which MCOs are present in a particular market and further, which MCOs and hospitals have contracts, by necessity we assume that this information does not affect the bargaining strategy between any MCO/hospital pair.

\textsuperscript{97}This assumption is clearly violated when a contract also requires that the hospital does not give larger discounts to other MCOs as in the case of Michigan Blue Cross, which has been sued by the U.S. Justice Department for antitrust violations (http://www.reuters.com/article/idUSN1827666920101018).

\textsuperscript{98}Brooks et al. (1997) are the first to incorporate the Nash bargaining game as the pricing mechanism and implicitly assume that there is only one hospital and one MCO in the market so that patients must be tied to the MCO. In consequence, if the two fail to reach an agreement then instead of seeking treatment at a different in-network hospital, patients will continue to visit the hospital and the insurer will have to pay list prices. Capps et al. (2003) utilize the bargaining outcome of Brooks et al. (1997).
When a hospital can threaten only its removal from network $\mathcal{M}$ the difference in the profits for MCO $m$ and hospital $h$ can thus be expressed as

$$\Delta\Pi_m(p_{hm}) = \Pi_m(\mathcal{M}) - \Pi_m(\mathcal{M} \setminus h) = \Delta W_h(\mathcal{M}) - \Delta h R(P_{\mathcal{M}}), \quad (3.6a)$$

$$\Delta\Pi_h(p_{hm}) = \Pi_h(\mathcal{H}) - \Pi_h(\mathcal{H} \setminus m) = p_{hm} D_{\mathcal{M}}(h) - \Delta C_h(m), \quad (3.6b)$$

where $D_{\mathcal{M}}(h)$ is the *ex ante* expected demand for hospital $h$ from enrollees in MCO $m$’s network $\mathcal{M}$; \cite{99} $\Delta h R(P_{\mathcal{M}})$ is the difference in MCO $m$’s reimbursements to hospitals caused by removing hospital $h$ and $m$’s enrollees reallocating themselves to the remaining hospitals when the equilibrium price matrix is $P_{\mathcal{M}}$; \cite{100} and $\Delta C_h(m)$ is the expected change in cost to hospital $h$ when it joins MCO $m$’s network and increases its demand. Alternatively, when a hospital can threaten removal of all of its system members the difference in profits for MCO $m$ is expressed as

$$\Delta\Pi_m(p_{hm}) = \Pi_m(\mathcal{M}) - \Pi_m(\mathcal{M} \setminus S_h) = \Delta W_h(\mathcal{M} | S_h) - \Delta S_h R(P_{\mathcal{M}}), \quad (3.7)$$

where $\Delta W_h(\mathcal{M} | S_h)$ is hospital $h$’s share of the incremental value to enrollees of adding system $S_h$ to MCO $m$’s network $\mathcal{M}$. \cite{101} We proceed with deriving the model.

\cite{99}By necessity we assume that revenue is linear with quantity due to insufficient data to estimate a revenue function.

\cite{100}Formally:

$$\Delta h R(P_{\mathcal{M}}) = \sum_{k \in \mathcal{M}} p_{km} D_{\mathcal{M}}(k) - \sum_{k \in \mathcal{M} \setminus h} p_{km} D_{\mathcal{M} \setminus h}(k).$$

\cite{101}Hospital $h$’s share is calculated by adding its incremental value to the difference in the incremental value of the entire system and the sum of each individual system hospital’s incremental value multiplied by the fraction hospital $h$’s incremental value represents of the sum of each individual system hospital’s incremental value; i.e.,:

$$\Delta W_h(\mathcal{M} | S_h) = \Delta W_h(\mathcal{M}) + \left[ \Delta W_{S_h}(\mathcal{M}) - \sum_{k \in S_h} \Delta W_k(\mathcal{M}) \right] \times \frac{\Delta W_h(\mathcal{M})}{\sum_{k \in S_h} \Delta W_k(\mathcal{M})}.$$
under the assumption that when an MCO and hospital fail to reach an agreement, only the hospital is removed from the network. The derivation follows similarly when the entire system is removed.

By plugging the profits into (3.5) and taking the first-order condition, the bargaining outcome can be expressed as

$$\Delta \Pi_h(p_{hm}) = \alpha_h \left[ \Delta W_m(h) - \Delta C_h(m) + \Delta_h R(P_M) - p_{hm} D_M(h) \right].$$  \hspace{1cm} (3.8)

The term in brackets on the RHS represents the total amount of surplus generated by the hospital and MCO successfully negotiating a contract. In this way, the observed profit for a given hospital is a function of the value the hospital brings to a MCO’s network, the hospital’s treatment costs, the reimbursement rates for alternative hospitals, and the hospital’s bargaining power $\alpha_h$. To simplify the notation somewhat, let

$$\Delta R_m(h) = \Delta_h R(P_M) - p_{hm} D_M(h)$$

$$= \sum_{k \in M \setminus h} p_{km} \left[ D_M(k) - D_M \setminus h(k) \right].$$

That is, $\Delta R_m(h)$ identifies the change in MCO $m$’s costs (reimbursements to hospitals) net the reimbursements to hospital $h$ when $h$ is in $m$’s network $M$. Using $R_m(h)$, (3.8) can be expressed as:

$$\Delta \Pi_h(p_{hm}) = \alpha_h \left[ \Delta W_m(h) - \Delta C_h(m) + \Delta R_m(h) \right].$$  \hspace{1cm} (3.9)

To further identify how hospital, system, and market characteristics affect the bargaining power of hospital $h$, the bargaining power $\alpha_h$ is parameterized as

$$\alpha_h \equiv \alpha_0 + \beta H_h + \delta S_h + \eta M_h,$$  \hspace{1cm} (3.10)
where $H_h$ are hospital $h$’s characteristics that affect bargaining power (e.g., ownership type, physician arrangements, teaching status, system membership), $S_h$ are system characteristics that affect bargaining power (e.g., the concentration of the system within a single patient market, whether the system contains a teaching hospital), and $M_h$ are market characteristics for $h$’s market that affect bargaining power (e.g., the concentration of MCOs).

We have two issues that prevent us from estimating (3.9). First, our data do not allow us to observe the negotiated division of surplus between hospital-MCO pairs, and second, our patient choice data does not include out-of-pocket costs, preventing us from estimating the value for $\gamma_p$ in the demand specification and using that estimate to calculate the value of adding a hospital to an MCO in dollars ($\Delta W_h(M) = \gamma_p^{-1} \Delta V_h(M)$). Consequently, the model that our data permit us to estimate is

$$\Delta \Pi_{h,t}(p_{hm}) = (\alpha_0 + \beta H_h + \delta S_h + \eta M_h) \times [\gamma_p^{-1} \Delta V_{h,t}(M) - \Delta C_{h,t}(m) + \Delta R_{m,t}(h)] + \epsilon_{h,t},$$

where $t$ indexes time, $\Delta \Pi_{h,t}(p_{hm})$ is the average change in profit from contracting with an MCO, $\Delta V_{h,t}(M)$ is the average change in *ex ante* value (in *utils*) for adding hospital $h$ to MCO $m$’s network $M$, $\Delta C_{h,t}(m)$ is hospital $h$’s average change in cost for treating MCO $m$’s enrollees, $\Delta R_{m,t}(h)$ is the average change in expenditures by MCO $m$ for those patients that would choose hospital $h$ if it were available, but instead choose other hospitals within network $M \setminus h$, and $\epsilon_{h,t}$ is an i.i.d., zero mean error term, all at time $t$. The average profit, value, and cost represent the profit, value, and cost changes due to the average change in managed care patients for a given hospital.
and year. For example, if five MCOs have an equal share of the market, then the average consists of the change in profit, value, and cost using 20% of the total number of managed care patients predicted to visit hospital $h$. The dollar value of a *util* is free to change across years, reflecting any possible changes in tastes.\(^\text{102}\)

The parameters within the bargaining power term (i.e. the term in parenthesis) are identified off of variation across hospitals in the amount of profits earned relative to the size of the surplus generated by the hospital. The utils-to-dollars conversion factor $\gamma_{p}^{-1}$ is identified separately from the other parameters because the surplus generated by a hospital depends not only on the utility the hospital provides to patients but also on the hospitals costs of treating those patients.

Despite representing an average measure, Eq (3.11) generates coefficient estimates that are easily interpreted. For example, if we were to include a dummy indicating whether a hospital belonged to a system then that estimate identifies the average amount system membership contributes to a hospital’s bargaining power over a similar hospital which is not in a system; accounting for the fact that system membership may increase (or decrease) the available surplus due to the consumers’ willingness-to-pay for the system membership.

\(^{102}\)It should be noted that in Eq. (3.11) it is not entirely accurate to refer to $\gamma_{p}^{-1}$ as the conversion factor for *utils*-to-dollars because the parameter also absorbs the percentage that premiums are marked-down from consumers’ willingness-to-pay, which is a function of the competitiveness in the insurance market. We find that $\gamma_{p,t}^{-1}$ is different between rural and urban markets, though controlling for this difference has no impact on the point estimates for the bargaining power parameters. We find no statistically significant difference in $\gamma_{p,t}^{-1}$ between regions such as Southern California compared to Middle and Northern California.
3.4 Data

The data come from several sources. Hospital characteristics come from both the American Hospital Association’s (AHA) 2008 Annual Survey of Hospitals, the AHA’s Hospital guide for 2007 and 2008, and the California Office of Statewide Health Planning and Development (OSHPD) Financial Disclosure Reports for 2001 through 2009. Financial data also come from the OSHPD Financial Disclosure Reports and discharge data come from the OSHPD Patient Discharge Reports for 2007 and 2008. The AHA survey has data for 399 of the 459 California hospitals. Of these, 29 hospitals were removed from the sample because no suitable match could be made between the AHA and OSHPD data sets. The hospital characteristics include properties of the hospital such as its ownership type (government, profit, and non-profit), teaching status, and system and network membership, as well as dummy variables for the services the hospital offers. Table 3.7 provides frequencies for various hospital attributes and Table 3.2 provides summary statistics for the operating characteristics of the private hospitals.

The OSHPD Financial Reports include data on each hospital’s total operating costs, gross revenues by payer (Medicare, Private, Medicaid, Worker’s Comp., and Self-Pay), and net revenues, as well as data on the number of in-patient days, discharges, and out-patient visits by payer. Total operating expenses, in-patient days, and out-patient visits are used to estimate the hospital cost functions. We use the

\[103\]

For excluded hospitals, neither the hospitals’ name or address represented a match. In some instances a hospital’s name and address differ between the two data sources; however, the addresses are for the same campus, thus identify the same hospital.
net and gross revenues to generate a rough estimate of the average revenue per in-patient day. Net revenues are provided as an aggregate by payer for both in- and out-patient services whereas gross revenues are provided by payer, subdivided into in-patient and out-patient totals. To estimate the average revenue per in-patient day for managed care patients we first calculate the deduction ratio by payor by dividing the total net revenues by the total gross revenues for each payer. Then, to account for revenue differences for patients admitted through the emergency room versus those with scheduled appointments we multiply the deduction ratio by the total charges for payer and admittance type divided by the total number of in-patient days by payer and admittance type.\textsuperscript{104} Our focus is on Medicare and privately insured patients only so we generate a total of four revenue estimates for each hospital (2 payers and 2 types of admittance).

The OSHPD Patient Discharge Reports contain 4,012,774 and 4,017,998 discharges for all acute care hospitals in the state of California for 2007 and 2008, respectively. The number of discharges is reduced based on several factors. First, many discharges are eliminated based on the hospital to which they belong. Specifically, only discharges for hospitals categorized as \textit{Comparable} by the OSHPD are

\textsuperscript{104}For example, the revenue per day for a patient in a managed care, Medicare (MC-M) plan who was admitted through the emergency room (MC-M-ER) is calculated as:

\[
\text{Avg. Rev./Day} = \frac{\text{Net Rev. for MC-M}}{\text{Gross Rev. I.P. for MC-M} + \text{Gross Rev. O.P. for MC-M}} \times \frac{\text{Total Charges for MC-M-ER}}{\text{Total I.P. days for MC-M-ER}}
\]
considered.\footnote{From the OSHPD documentation: *Comparable* includes hospitals whose data and operating characteristics are comparable with other hospitals} This excludes patients from Kaiser hospitals (27 hospitals),\footnote{Excluding Kaiser hospitals does not affect the analysis because Kaiser is an HMO providing both the insurance and hospital services and we are examining the bargaining power of hospitals conditional on patients enrolling with a non-Kaiser insurer.} state hospitals providing care to the mentally disordered and developmentally challenged (7 hospitals), Shriner Hospitals for crippled children (2 hospitals), and long-term care hospitals (2 hospitals).

We further eliminate discharges from the data based on some characteristics of the patient and discharge. Only acute care discharges are considered, eliminating all intermediate and skilled nursing, psychiatric, chemical dependency and recovery, and physical rehabilitation care. We exclude all discharges for infants under 24 hours of age and other unknown types of admission. All discharges for patients not originating from the state of California as well as those discharges originating from hospitals over 90 minutes from the patient’s zip code are also eliminated from the sample.\footnote{18,679 or slightly less than 2 percent of the discharges in the sample are from admissions scheduled in advance for hospitals over 90 minutes away.} Finally, all discharges for patients with a masked age category are not included.

Although we would ideally like to identify how each DRG affects hospital demand we do not have a sufficient number of observations for each DRG to produce a precise estimate of demand so we instead categorize a patient’s illness by the Major Diagnostic Category (MDC). Of the 25 MDCs, we eliminated all MDCs accounting for less than one half of a percent of the total discharges, MDCs for which there may be significant non-hospital competition (e.g., mental diseases and disorders), as well
as the discharges related to neonatal care leaving 15 MDCs. Finally, only patients insured by either Medicare and under 75 years of age or privately insured are considered. Medicaid patients may have a restricted set of hospitals from which they can choose and worker’s compensation and self-pay patients may also have preferences differing significantly from the privately insured population. The final sample contains 2,027,323 discharges.

We also interact the MDCs with many of the relevant characteristics of the hospital. For example we interact a diagnosis of child birth with a dummy indicating if the hospital has a birthing room. We include seventeen such interactions, which are also interacted with travel time. Some hospital characteristics are also interacted with the patient’s age category when we have reason to believe that service may be particularly important for a patient of a particular age; e.g., neurological services are found to be particularly valuable for patients over 65 years of age. A total of 155 covariates are included in the demand estimation.

Descriptive statistics for the included discharges are reported in Table 3.5. The characteristics of the discharges are divided into four categories. The category “Insurer” indicates if the discharge is for a privately insured patient or a Medicare patient. The category “Type” indicates if the patient’s insurance represents managed care or indemnity, fee-for-service. The category “Diagnosis” has only one item, the patient’s length of stay, which is a measure of the severity of a patient’s illness. The category

108 The six MDCs eliminated include: eye diseases and disorders, mental diseases and disorders, poisonings and toxic effects of drugs, burns, multiple significant trauma, and human immunodeficiency virus infections.
“Diagnostic Category” provides the shares for each of the 15 included MDCs. The final category “Patient” identifies the patients’ characteristics and include the patients’ gender, age, race, number of hospitals in the patients’ choice sets, travel time to the chosen hospital, the travel time to the other hospitals in the patient’s choice set, whether or not the patient comes from a rural location, and the average income for the patient’s zip code.

3.5 Results

3.5.1 Demand

The estimates from our demand model are primarily used to calculate differences in the willingness-to-pay for different hospitals and to simulate consumer choice under different choice sets. In general we will refer to quality as any characteristic that makes a hospital more attractive to a patient. Table 3.7 provides coefficient estimates for many of the patient and hospital characteristics for each year. Demand is estimated using only patients enrolled in traditional insurance giving them unconstrained choice sets. All estimates are significant at the 1 percent level. Most of the estimates have the expected sign, for example teaching hospitals have higher quality and patients are willing to travel further to visit one. The nurse-to-bed ratio is also found to improve a hospital’s quality and again patients are willing to travel further to visit a hospital with a higher ratio. Wealthier patients are willing to travel further than poorer

---

109 Travel time is calculated using the Google Maps API and is calculated as the travel time from the patient’s home zip code centroid to the hospital taking into account traffic patterns, speed limits, and stop lights. For a couple rural zip codes there is no travel route from the zip code centroid because it is a desert or national or state park. In these cases the center of the nearest small town within the zip code is used and in all cases there is only one such town.

110 This could be actual quality of care, perceived quality of care, or some other amenity.
patients. The preference for travel time by wealth may indicate that poorer patients have less options due to public transportation routes. The estimates are mixed with respect to length of stay interacted with travel time having a negative coefficient in 2007 and a positive coefficient in 2008. Having a longer length of stay may be more inconvenient in the sense that the patient can expect less visits from friends and family if the hospital is further away, but supports the conjecture that length of stay is a good proxy for case severity and patients are willing to travel further for more severe conditions. The majority of MDC interactions are statistically significant at the 1 percent level as well. Finally, the highest probabilities within a choice set correspond with the chosen hospital for 38% of the discharges, which is a good prediction rate for a logit demand model.

Table 3.8 provides estimates for the change in value in \textit{utils} for various types of hospital systems. Because system memberships are interacted with travel time there is no single coefficient estimate that identifies the differential value of system membership. Instead, we present the estimated difference in value for a system hospital relative to a non-system hospital for the median patient for that MDC, \textit{ceteris paribus}.\textsuperscript{111} For example, the first row of Table 3.8 identifies the difference in value for a patient diagnosed with an illness of the nervous system (MDC 1). The median distance from the chosen hospital for such a patient is 16 minutes. The first column identifies the difference in value for a not-for-profit hospital from not being in a system to being in a local system that does not cross state lines, and it is negative (−.0922).

\textsuperscript{111}In theory we could use the estimate for $\gamma_p$ to express the difference in dollars; however, because $\gamma_p$ is absorbing the difference between the premium and the willingness-to-pay as well, it is an underestimate of the \textit{utils}-to-dollar conversion.
Similarly the second column identifies the difference in value for a not-for-profit hospital from not being in a system to being in a system that crosses state lines,\textsuperscript{112} and is also positive (.1198).\textsuperscript{113}

The only consistently positive difference in value comes from systems that are national in scope, independent of whether the hospital is not-for-profit or for-profit. Interpreting the point estimates is somewhat challenging given the nebulous units that define a \emph{util}; however, we can gain some perspective by comparing the difference in \emph{utils} from system membership with the difference in \emph{utils} from a teaching hospital. For example, for a median patient 16 minutes from the hospital, teaching status adds 0.3479 \emph{utils} of value over an identical non-teaching hospital in 2008. In comparison, a patient visiting a hospital for child birth (MDC 14) values a for-profit hospital by 0.6130 \emph{utils} more if that hospital is in a national system, than if it is not, indicating that the value added by system membership can be fairly significant depending on the diagnosis. The largest positive difference on a hospital’s value with system membership is for pregnancy and childbirth for a for-profit hospital—the most common diagnostic category—and for diseases of the musculoskeletal system (MDC 8) for a not-for-profit hospital (though the value-added for child birth is also quite high for a not-for-profit hospital). The difference is not as large for the second most frequent diagnostic category—the circulatory system—which includes DRGs for chest pain and

\textsuperscript{112}For clarity we refer to systems that span state lines as a \textit{National} system. Many of these systems are truly national in scope, but several are more regional concentrating in the west coast or the south western states.

\textsuperscript{113}We tried different demand specifications having various categorizations of system status. For example we tried categorizing systems as either single-hospital or multi-hospital systems, however, we found that most of the single-hospital systems in California belonged to national chains so were in fact multi-hospital systems. Reflecting this, there was no real difference between types of system as there is when we bifurcate systems based on their scope.
myocardial infarction. This may reflect the fact that the MDC contains a mix of
diagnoses having enormous variability in expected costs whereas the MDC for child
birth and pregnancy contains very few DRGs with less variability in expected costs.
Or it may be the case that hospitals compete more over cardiac care than labor and
delivery.

3.5.2 Cost

We tried several variations of the cost model by including system fixed effects,
and interacting system membership with hospitals’ outputs. All of the cost models
successfully generate predicted costs that are exceptionally close to the realized costs
for a given mix of inputs and outputs. However, because our purpose for estimating a
hospital’s cost function is to use that function to estimate the marginal cost of treating
an MCO’s enrollees, we picked the specifications balancing a tighter predicted fit with
higher precision on the various hospital outputs. Coefficient estimates are reported
in Table 3.9.

Inpatient days are grouped by payer and include Medicare patients, the privately
insured, and a catch-all for all other payers such as MediCal, self-pay, and worker’s
compensation. The coefficients of interest for the purposes of estimating marginal
costs of treating managed care patients are the coefficients associated with discharges
from Medicare and privately insured patients. All of the hospital outputs are also in-
teracted with a for-profit dummy to account for any marginal cost differences between
the type of hospital. The coefficients on in-patient days for Medicare and privately
insured patients are all significant at the 1% level and have mixed sign between the first- and second-order terms.

All of the cost models include fixed effects for time and market area, indicators for profit status, teaching status, rural status, the presence of a PET and/or MRI machine as well as terms for outpatient visits, hospital beds, number of staff, and percentage of patients which are insured by Medicare. Cost model (C1) is the most basic in that it does not include any controls for system membership. Cost models (C2) and (C3) include a dummy for system membership and (C3) additionally includes interactions of teaching status with outputs. Cost model (C4) includes interaction terms for system membership with outputs to account for any efficiency changes that are either realized through system membership, or lead to system membership. Finally cost model (C5) includes interactions of teaching and system membership with the hospital’s outputs.

The average cost of an in-patient day is 9.76% higher at a not-for-profit hospital than at a for-profit hospital. This suggests that either for-profit hospitals are more efficient or admit lower risk patients on average. The average cost of an in-patient day only highlights part of the difference in costs, though. To see how else the two types of hospital differ we used the estimated cost functions to calculate the difference in total costs at the sample means. Table 3.11 reports these cost differences. Each row represents the cost differential based on applying the sample means to the estimated cost functions. The second column reports the overall difference in cost for a not-for-profit and for-profit hospital, the third column reports the difference in costs that is attributable to the difference in characteristics of the hospitals (e.g., the differences
in patient volumes, nurse-to-bed ratios, etc.), and the fourth column reports the
difference in costs attributable to the difference in the parameter estimates (i.e., the
difference in the hospital’s technologies or cost functions). The results show that
not-for-profit hospitals have substantially higher costs than for-profits with an overall
difference ranging from 253% in cost specification (C5) to 321% in cost specification
(C2). The differential in costs is largely driven by a significantly larger patient volume
at not-for-profit hospitals. For instance, the average number of out-patient visits at a
not-for-profit hospital is 166,310 compared to 48,508 at a for-profit, and the average
number of discharges are 11,525 at a not-for-profit, compared to 7,582 at a for-profit
hospital.¹¹⁴

We also find that the average cost of an in-patient day is 8.5% higher at a system
hospital than at a non-system hospital. As with not-for-profit hospitals, this may
indicate that system hospitals are less efficient, or that system hospitals have a more
severe patient case mix. We find that there is no statistically significant difference
between the propensity of a for-profit compared to not-for-profit hospital to be in a
system with for-profit hospitals representing about 25 percent of the system hospitals
in the sample. We also applied the sample means for the two types of hospitals
to the estimated cost functions and the cost differences are reported in Table 3.7.
Though not as dramatic of a difference as found with for-profit and not-for-profit
hospitals, on average system hospitals have higher overall costs ranging from 27% in
cost specification (C5) to 41% in cost specification (C2). About 10 to 20 percent of
the cost difference is due to a difference in the hospitals’ technologies.

¹¹⁴Both patient-volume differences are statistically different at the $p = .01$ level.
3.5.3 Bargaining Power

Using the demand and cost estimates from Sections 3.5.1 and 3.5.2 we calculate each hospital’s expected costs and profits from treating managed care patients and use that to estimate the percentage of the available surplus received by the firm. Recall from Eq. (3.9) that the surplus over which each MCO-hospital pair bargain is based on the expected number of managed care patients that belong to the specific MCO with which the hospital is negotiating. Because a hospital’s expected difference in cost from treating 500 additional patients when it is currently treating 1,000 patients enrolled with other insurers is different from when it is currently treating 2,000 patients, we must be careful to identify how many patients come from a representative MCO and how many patients come from the other MCOs.

Unfortunately we cannot observe with which MCOs a hospital has contracts, nor can we observe the number of contracts.\textsuperscript{115} Despite the market having 49 insurance companies with managed care plans in 2008, not counting the integrated HMO Kaiser Permanente (California Department of Managed Health Care), Figure 3.1 indicates that the market is predominately covered by a small number of insurers. For example the largest insurer, Blue Cross of California, has a 26% market share, but the sixth largest insurer, Aetna, has only a 3% market share. Table 3.4 lists the top ten insurers and their market share, showing that the first five insurers account for about 80% of the total market.

\textsuperscript{115}The AHA annual survey actually asks hospitals to report the number of HMO and PPO contracts they have, but only 137—accounting for about a third—of the private hospitals reported these numbers.
Given the rapid decline in market shares and the fact that the top five account for a significant share of the market, we assume that 15% of the total number of managed care patients that are predicted to visit a hospital come from a single representative MCO and calculate the change in a hospital’s cost and revenue as the difference between the cost in treating all of the managed care patients predicted to visit the hospital and the cost of treating 85 percent of the predicted managed care patients. Furthermore, to keep us in the area of the cost function containing the data, we difference the predicted managed-care patients from the sum of the total number of managed care patients that are predicted to visit the hospital plus all of the non-managed care patients who we observe visiting the hospital plus all of the managed care patients that we observe visiting the hospital but are not a part of the out-of-sample demand data. To test the robustness of the results to the sample proportion we estimate the model using various proportions of the predicted managed care patients and these results are reported in Table 3.16. Similarly, the expenditures for an MCO are calculated by taking a 15% sample of the managed care patients predicted to visit a specific hospital and using the demand estimation to simulate where those patients would go if that hospital is removed from their choice set.

Table 3.13 reports the bargaining estimates for one bargaining specification utilizing data generated from the five cost specifications reported in Table 3.9 grouped by whether a hospital can threaten its removal alone (Hosp.), or the removal of its entire system (Sys.). All standard errors are clustered by hospital and following Murphy and Topel (1985), adjusted to account for the data generated from the first-stage cost and demand estimate. The regression specification includes several hospital and
market characteristics. Most of the signs are as would be expected with a couple notable exceptions which we will discuss in turn. The first row should be interpreted as the regression constant and is the base bargaining power for a hospital. The overall bargaining power for a hospital is found by adding the coefficient estimates for the hospital and market characteristics to this base bargaining power. There is slight variation with the estimates.

Previous research such as Cuellar and Gertler (2006) has found evidence that hospital arrangements with physician groups lead to higher reimbursement rates, with evidence of increased quality in some cases. To control for how the various forms of vertical integration between physicians and hospitals may influence bargaining power we included a set of dummies indicating whether or not the hospital is affiliated with a Closed Physician-Hospital Organization (CPHO), Open Physician-Hospital Organization (OPHO), Independent Practice Association (IPA), or a Group Practice Without Walls (GPWW). We found no correlation between CPHOs, OPHOs or IPAs and hospital bargaining power. Affiliation with a Group Practice Without Walls is found to be correlated with a relatively large increase in a hospital’s bargaining power. Given the small proportion of hospitals which report affiliation with a GPWW as well as the large number of hospitals which fail to report whether or not they have affiliations, these results should be interpreted with caution. This result, however, suggests the need for further research to identify the effect physician arrangements have on a hospital’s bargaining power, particularly given the recent emphasis policy makers have put on Accountable Care Organizations and strengthening the vertical relationship between physicians and hospitals.
Teaching status is found to be correlated with an increase in bargaining power, however, the estimates are not significantly different from zero in any model. Intuitively a teaching hospital should secure a higher reimbursement rate, and the positive coefficient estimates in the demand estimation show that consumers assign a significant value to teaching hospitals. Nevertheless, having controlled for the difference in consumer value derived from teaching status, it is not clear that a teaching hospital should have more bargaining power.\footnote{We spoke with an executive from a major national MCO, and he indicated that teaching hospitals are generally provided a higher reimbursement rate to compensate for the hospital’s higher costs of treatment, but he did not think that the higher reimbursement rate completely covered the higher costs, suggesting little to no increase in bargaining power.}

Rural status is associated with weaker bargaining power, though the estimates are only consistently significant in the Hosp. model due to smaller point estimates in the Sys. model. \textit{A priori}, the bargaining power of a rural hospital could be higher or lower relative to an urban hospital. First, because a rural hospital is more likely to have monopoly power, the hospital could conceivably exercise that market power to extract most of the surplus. Because a rural hospital is likely to represent only a small percentage of an MCO’s enrollment base, however, the bargaining power can easily tip in the direction of the MCOs if they do not view the hospital’s patient market as critical to their business strategy. We also cannot rule out the possibility that the marginal costs for a rural hospital have been over-estimated resulting in an under-estimation of the rural hospital’s bargaining power. Given the paucity of rural hospitals we were not able to estimate separate marginal costs with any precision.

To test whether or not larger hospitals have more bargaining power we include the number of beds and the number of beds squared. The results have mixed signs and
are not statistically different from zero providing no evidence that larger hospitals have more bargaining power. We also tried using dummies indicating if a hospital is in the 50-75th percentile, 75th to 90th percentile, and higher than the 90th percentile in number of beds and again found no evidence that larger hospitals have more bargaining power. This does not mean that larger hospitals are not able to exercise market power to extract relatively more profit, though. Given the concavity in the incremental value of adding a hospital to an MCO’s network with respect to number of beds, a larger hospital should generate proportionately more surplus, thus profit, than a smaller hospital (Chipty and Snyder, 1999; Horn and Wolinsky, 1988; Inderst and Wey, 2003; Stole and Zwiebel, 1996).117 However, conditional on the difference in surplus, there is no reason that a larger hospital should be able to extract a higher proportion of the surplus than a smaller hospital.

To control for the market power that a hospital may have within its patient market we included the market share of a hospital within a 10 mile radius.118 Surprisingly the coefficient estimate was low and statistically insignificant, however, when we include the square of the market share the estimates are positive and significant for

117 To see this, recall that the marginal value of adding a hospital to a network is given as \( \Delta V = N \int \ln \left( \frac{1}{1 - s} \right) dF \). Letting \( Y \) represent the number of patients who receive treatment at the added hospital, which is strongly correlated with hospital size, and \( X \) represent the number of other patients enrolled with the MCO, the marginal value can be rewritten as \( \Delta V = N \int \ln \left( \frac{X + Y}{X} \right) dF \), which is concave in \( Y \).

118 We limited the market share to a 10 mile radius because, on average, 72% of a hospital’s managed care population originate from within 10 miles of the hospital whereas only 18% originate from a distance of 10 to 20 miles, and only 8% from a distance of 20 to 40 miles. Therefore, if we extend the radius too far, then all hospitals will appear to have fairly low market shares and, if we crop the radius in too tight, then hospitals will appear to have very high market shares creating too little variation between hospitals.
market share, and negative and significant for market share squared indicating that bargaining power is concave, increasing with a hospital’s market share.

To compare with the earlier literature on hospital prices and system membership we include a simple dummy indicating whether a hospital is in a system. Correlating with the findings of earlier papers (Ho, 2009; Melnick and Keeler, 2007), which find that system hospitals have higher reimbursement rates independent of the market power generated from increasing concentration within the local patient market, system membership is also associated with higher bargaining power using most of the cost specifications. Cost specifications (C1) in either the Hosp. or Sys. model and (C3) did not generate a statistically significant estimate in the Sys. model, however, they did generate point estimates of similar magnitude to cost specifications (C2), (C4), and (C5) which are all statistically significant at the 5 to 10 percent levels. As expected, the point estimates are all higher in the Hosp. models as bargaining power absorbs all of the relationship higher profits and system membership as there is no change in the MCO’s disagreement point. The estimates for the Sys. models provide clear evidence that, even after controlling for the added value that system membership may provide patients as well as the system’s market power, system membership is associated with a bump in bargaining power, *ceteris paribus*. In Section 3.5.3 we include several system characteristics to try to identify what specific characteristics of system membership may result in the bargaining power differential.

To control for the concentration of MCOs within a hospital’s market we included a measure of market share for Blue Cross/Blue Shield of California (BCBS), the largest MCO operating in the state. The market share is measured as the percentage
of patients within a 10 mile radius of a hospital which were insured by BCBS. The presumption is that a higher share indicates a more concentrated MCO market. As expected the sign is negative and generally statistically significant.

Finally, Table 3.13 also reports the estimates for the utils-to-dollar conversion factors, $\gamma_p$ and $\Delta \gamma_p(2007)$. As the differences in the demand estimates across 2007 to 2008 indicate, there is considerable difference in the conversion factors between years. Moreover, there is a slight inverse relationship between the conversion factor and the base bargaining power. For example, for both the Hosp. and Sys. models cost specifications (C2) and (C5) produced the highest and lowest base bargaining power estimates, respectively, while also producing the lowest and highest conversion factors.

**Hospital Systems and Bargaining Power**

To better identify the characteristics of hospital systems that are associated with bargaining power we try several different regression specifications which include different measures of system size and market coverage. We again perform the analysis assuming a system hospital can threaten only its removal from an MCO’s network (Hosp.) and assuming a system hospital can threaten removal of all of its system members (Sys.). The results of these regressions are reported in Tables 3.14 and 3.15, respectively. All of the regressions are performed using cost specification (C5)—the most inclusive specification—and estimated using a 15% sample of the predicted managed care patients. All of the specifications include the same regressors included in Table 3.13; however, only the base bargaining power, Blue Cross/Blue Share and hospital market shares are reported.
Each specification includes an indicator identifying if the hospital is in a system. Because For-profit (FP) and Not-for-profit (NFP) hospitals appear to utilize their system differently we interacted system membership along with the other system-specific covariates with the profit-status of the hospital. The estimates in (A) and (B) indicate that, on average, both FP and NFP system hospitals extract more of the surplus than non-system hospitals, however, the difference is smaller and not statistically significant in the Sys. model. In general, the estimated bargaining power advantage of system hospitals should be lower in the Sys. model because some of the increased hospital profit is attributed to the market power effect of the system, while in the Hosp. model all of the increased hospital profit must be explained through a higher bargaining power parameter.

Because a system and MCO likely have multi-market contact when systems are present in multiple patient markets, there is a possibility that systems are able to leverage their strength in one market to extract more of the surplus in another; i.e., use a tying arrangement to increase bargaining power. To see if there is any evidence that systems that span patient markets also have higher bargaining power we included several covariates related to multiple market systems. First we include a dummy indicating whether or not a system includes a teaching hospital, which are

\[119\] For example, Whinston (1990) shows that if the market structure for a tied good is oligopolistic and exhibits scale economies, then tying can be profitable for a monopoly hospital via market foreclosure. Carlton and Waldman (2002); Choi (1996, 2004); Choi and Stefanadis (2006) all examine different factors that affect the suitability of using tying arrangements. These studies focus on some means of market foreclosure as the mechanism leading to a successful tying arrangement. For example Carlton and Waldman (2002) examine how a hospital with a monopoly in one market can use tying to prevent a competitor with a superior complementary product from entering the monopolized market and Choi (1996, 2004); Choi and Stefanadis (2006) all examine how tying arrangements can be used to reduce an entrant’s R&D incentives and thus the success of entry.
high valued hospitals that MCOs generally do not want to exclude. The coefficient estimate is both close to 0 and statistically insignificant. Second, we include a dummy, again interacted with profit-status, indicating whether or not a hospital’s system operates in more than one patient market—e.g., the system has hospitals in both the Los Angeles and San Francisco areas. Interestingly, we find that FP hospitals belonging to a multi-market system do have higher bargaining power than FP system hospitals that do not belong to multi-market systems, whereas NFP hospitals do not. The fact that multi-market system hospitals extract more of the rents could suggest a transaction-costs/scale economies story; though, this explanation is tempered by the fact that NFP hospitals do not enjoy the same increase in rents. Finally, we also tried including the highest market share of one of the hospitals in the system and, similar to having a teaching hospital in the system, did not find any evidence that hospitals could leverage strength in one market to extract rents in another.

Given the amount of bargaining power associated with the market share of a hospital we included the market share of a hospital’s system members within a 10 mile radius of that hospital. As with own market share, the estimate is positive and statistically significant for FP hospitals in either the Hosp. or Sys. models, further indicating that market concentration within a local patient market endows a hospital with significant bargaining power vis-á-vis an MCO.\textsuperscript{120} For example, a one standard deviation change in own-system market share in the patient market surrounding a hospital results in a 11.03 (D) to 12.55 (F) percentage point increase in the share of

\textsuperscript{120}We also tried including system share squared, but this value was highly statistically insignificant in contrast to own market share so it was removed from the final specification. This difference is attributable to the much lower variance and range of system market shares and the fact that higher same system shares is not correlated with a rural market.
the surplus extracted by the hospital as profit in the Sys. model. Again there is no association between system market share and bargaining power for NFP hospitals.

Remarkably 18 percent of the system hospitals have a fellow system hospital within 2.5 miles. We hypothesize that hospitals within such close proximity have considerable bargaining power in their local vicinity—bargaining power that is not captured by the system share—therefore we include a dummy indicating if a hospital has a partner within this distance.\textsuperscript{121} The estimate is positive and statistically significant for FP hospitals in both models and positive for NFP hospitals, but statistically significant only in the Hosp. model. One reason we observe such a significant increase in bargaining power is because within California, the presence of a close partner is frequently associated with substantial market concentration. For example AHMC Healthcare—a for-profit system—has 5 hospitals clustered around the Monterey Park neighborhood of Los Angeles and Citrus Valley Health Partners—a not-for-profit system—has 3 hospitals clustered around Covina exerting significant monopoly power on those neighborhoods.

We also tried using several measures of the size of a system using the total number of beds in the system and the total number of beds squared, total number of member hospitals and the total number of member hospitals squared, as well as dummy indicators for different measures of size. The results were of mixed sign, but never statistically significant providing little insight into what, if any, effect system size has on bargaining power, though most systems in our data have at least three members so it may be likely that the marginal effect of adding more members drops rapidly.

\textsuperscript{121} Including a dummy instead of system market share helps us to avoid multicollinearity problems with the 10 mile system share.
conditional on how adding additional members affects market shares. Moreover measures such as system market share, the presence of a close partner, and belonging to a multi-market system are all highly correlated with system size so are likely picking up any differences in bargaining power associated with larger systems.

Lastly, to make sure that the largest systems in our data are not driving some of these results, we include indicators for those systems: Catholic Healthcare West, Adventist Health, and Sutter Health Including dummies for member hospitals of these three systems had virtually no affect on the point estimates for the other parameters. Furthermore, all of the system dummies have large standard errors suggesting significant variation in the bargaining power of individual hospitals within the systems.

3.5.4 Predicted Managed Care Patients and Bargaining Power

Before concluding this section we revisit the assumption regarding the average proportion of managed care patients that can be attributed to a single MCO. Instead of estimating a hospital’s bargaining power utilizing different cost specifications, Table 3.16 reports regression results where different proportions of the predicted managed care patients are used to estimate the hospitals’ change in costs and revenues, and the MCOs’ change in expenditures when a hospital is not in its network. All of the regressions utilize cost specification (C5) as the hospitals’ cost function. As with the previous regressions, the bargaining regression is run twice for each proportion (50%, 25%, 15%, 10%, and 1%) where the columns Hosp. refers to the scenario in which a hospital can threaten only its removal and Sys. refers to the scenario in which a
hospital can threaten the removal of all of its system partners. The middle group (15%) corresponds with the results in the last group (C5) reported in Table 3.13.

There are no notable deviations in the estimates in for the 50%, 25%, or 10% groups from the 15% group used in the main analysis. The fact that there is not much variation between estimates based upon the proportion of predicted managed care patients is reassuring in that the results are not driven by the choice of what proportion of patients to use for a representative MCO. At the extreme of 1% we do observe some deviations in the estimates from the other proportions. This difference suggests that the curvature of the cost function is important since for such a small fraction of patients the change in costs will be fairly linear for all of the hospitals. Given the few large market share MCOs operating in California the 1% sample is not appropriate for the analysis.

3.6 Market Power Versus Bargaining Power

There are a number of ways system hospitals can secure higher prices over non-system hospitals. For example, system hospitals can simply be more attractive to enrollees allowing MCOs to set higher premiums that get passed on to hospitals through higher reimbursements. On the supply-side there are two channels arising from hospitals’ relationships with MCOs that can result in higher reimbursement rates. First, by threatening to withdraw the entire system from an MCO’s network if a favorable reimbursement rate is not agreed to, a hospital lowers the MCO’s disagreement point, increasing the value of the hospital to the MCO. Some of the higher value is passed on to the hospital via higher reimbursement rates, generating
what we call the *market power* effect. However, higher hospital markups can also arise through a second supply side channel. Conditional on the willingness to pay by consumers to have access to the hospital and the increase in value to the MCO resulting from system membership, system hospitals may have greater bargaining power than non-system hospitals, allowing them to extract a larger share of the hospitals value to the MCO. Note that the market power effect only arises when system hospitals are able to threaten to withdraw other system hospitals from the MCO’s network, while the bargaining power effect can arise even in the Hosp. model where a hospital can only threaten its own removal from the network.

To identify the relative importance of the two supply-side channels we take hospitals’ average daily reimbursement rate (ADRR) and identify the amount of the difference between system hospitals’ ADRRs compared to a non-system hospitals’ ADRR that is attributable to differences in market power and to differences in bargaining power. Table 3.17 reports these decompositions. We report both percents and levels for the 75th, 50th, and 25th percentile hospitals, as well as the average percent and levels for each regression specification reported in Table 3.15.

Since each specification represents a slightly different version of the bargaining power equation, there is some variation in the size of the bargaining power effect across specifications. In contrast, the Market Power numbers do not change much because we only use one demand specification to estimate the incremental value of the hospital and its system to an MCO.\textsuperscript{122} Generally, about 80\% of the difference in reimbursement rates between system and non-system hospitals can be attributed

\textsuperscript{122}The variation that is present across specifications results from differences in the estimate for $\gamma_p^{-1}$. 

146
to differences in bargaining power. The 25th percentile values reveal that, at least in some specifications, there are some hospitals for which the estimated bargaining power effect is relatively small. Nevertheless, the results clearly reveal that differences in bargaining power across hospitals may be an important contributor to observed differences in reimbursement rates.

3.7 Conclusions

Our findings highlight an important, yet overlooked, channel through which hospital mergers may lead to higher reimbursement rates. While previous studies have shown that merging hospitals can increase their value to MCOs by reducing competition in the local market, our results suggest that mergers may also increase hospitals’ bargaining power, allowing them to extract a larger proportion of their value from MCOs. After controlling for the differences in value to MCOs as well as the costs to treat patients we find that system hospitals do extract a larger proportion of their value to MCOs than similar non-system hospitals. Indeed, we find that differences in bargaining power contribute more to the observed price differential between system and non-system hospitals than that created by differences in hospital market power.

We have uncovered a few system properties that are associated with higher bargaining power. Systems having a high market share within a local patient market have both higher value to the MCO and greater bargaining power. Moreover, hospitals having a close system partner have even higher bargaining power. The empirical evidence also suggests that hospitals—particularly for-profit hospitals—may be able to leverage their system affiliation, perhaps through multi-market contact with MCOs,
to extract more of the surplus beyond what can be done through local market power. For-profit hospitals that are members of a multi-market system are estimated to have higher bargaining power than comparable non-system hospitals. This finding raises important policy questions as current antitrust policy generally considers the local patient market as the relevant market definition when considering hospital mergers.

Finally, we find some evidence that hospitals affiliated with physician organizations, particularly GPWWs, are also associated with higher bargaining power. This is somewhat concerning given the recent move towards more vertical integration between physicians and hospitals through the formation of Accountable Care Organizations.

It should be noted that while our structural model relies on economic theory to identify the bargaining power of individual hospitals, the theoretical literature provides very little guidance in understanding why differences in bargaining power arise. Consequently, our analysis is only able to identify correlations between bargaining power and hospital characteristics (such as system membership). The findings are consistent with concerns commonly expressed in the industry, that hospital systems are able to leverage the size of the system both within and across markets by threatening significant disruption to a large number of the MCO’s patients (through a withdrawal from the MCO network). Nevertheless, future theoretical or empirical analysis of the mechanisms through which hospital systems or vertical integration may impact bargaining power is likely to be particularly valuable.
<table>
<thead>
<tr>
<th>Type</th>
<th>Characteristic</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>Government</td>
<td>.228</td>
</tr>
<tr>
<td></td>
<td>For-Profit</td>
<td>.210</td>
</tr>
<tr>
<td></td>
<td>Non-Profit</td>
<td>.558</td>
</tr>
<tr>
<td>Status</td>
<td>Teaching</td>
<td>.080</td>
</tr>
<tr>
<td></td>
<td>Rural</td>
<td>.196</td>
</tr>
<tr>
<td>Affiliation</td>
<td>System Member</td>
<td>.620</td>
</tr>
<tr>
<td></td>
<td>Network Member</td>
<td>.203</td>
</tr>
<tr>
<td>Control Conditional on System Membership</td>
<td>Government</td>
<td>.105</td>
</tr>
<tr>
<td></td>
<td>For-Profit</td>
<td>.269</td>
</tr>
<tr>
<td></td>
<td>Non-Profit</td>
<td>.626</td>
</tr>
<tr>
<td>Physician</td>
<td>CPHO</td>
<td>.014</td>
</tr>
<tr>
<td>Arrangement</td>
<td>OPHO</td>
<td>.022</td>
</tr>
<tr>
<td></td>
<td>GPWW</td>
<td>.025</td>
</tr>
<tr>
<td></td>
<td>IPA</td>
<td>.199</td>
</tr>
</tbody>
</table>

Table 3.1: All Hospital Characteristics \((N = 276)\)

<table>
<thead>
<tr>
<th>Item</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Market Share ((\leq 10 \text{ mi.}))</td>
<td>0.233</td>
<td>0.236</td>
<td>0.004</td>
<td>0.964</td>
</tr>
<tr>
<td>Licensed Beds</td>
<td>243</td>
<td>174</td>
<td>10</td>
<td>1024</td>
</tr>
<tr>
<td>Staffed Beds</td>
<td>190</td>
<td>139</td>
<td>10</td>
<td>902</td>
</tr>
<tr>
<td>Total Discharges</td>
<td>53,014</td>
<td>42,678</td>
<td>482</td>
<td>271,808</td>
</tr>
<tr>
<td>Percent Medicare</td>
<td>40.90</td>
<td>14.48</td>
<td>.51</td>
<td>82.5</td>
</tr>
<tr>
<td>Percent Discount from List Price</td>
<td>73.70</td>
<td>7.95</td>
<td>47.33</td>
<td>89.70</td>
</tr>
<tr>
<td>Daily Expenses ($)</td>
<td>39,657,173</td>
<td>38,405,615</td>
<td>98,769</td>
<td>254,812,268</td>
</tr>
<tr>
<td>Ancillary Expenses ($)</td>
<td>68,056,018</td>
<td>82,981,715</td>
<td>2,059,393</td>
<td>598,230,233</td>
</tr>
<tr>
<td>Total Operating Exp. ($)</td>
<td>200,630,443</td>
<td>233,382,012</td>
<td>5,654,633</td>
<td>1,747,307,004</td>
</tr>
</tbody>
</table>

Table 3.2: Private Hospital Operation Characteristics \((N = 196)\)
<table>
<thead>
<tr>
<th>System Size</th>
<th>System Name</th>
<th>Home State</th>
<th>System Size</th>
<th>System Name</th>
<th>Home State</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual</td>
<td>in Data</td>
<td></td>
<td>Actual</td>
<td>in Data</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>Success Healthcare</td>
<td>FL</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Trinity Health</td>
<td>MI</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>Department of the Army</td>
<td>VA</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Envision Hospital Corporation</td>
<td>AZ</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>Department of the Air Force</td>
<td>DC</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>MedCath, Inc</td>
<td>NC</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>Tahoe Forest Health System</td>
<td>CA</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Nexus Health Systems</td>
<td>TX</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Hospital Partners of America</td>
<td>NC</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Sisters of Charity</td>
<td>KS</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Banner Health</td>
<td>AZ</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>NorthBay Healthcare System</td>
<td>CA</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>Valley Health System</td>
<td>CA</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>Sun Healthcare Group</td>
<td>CA</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>HEALTHSOUTH Corp.</td>
<td>AL</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>Brim Healthcare, Inc</td>
<td>TN</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>Alta Healthcare System</td>
<td>CA</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>Shriners Hosps for Children</td>
<td>FL</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>Stanford Health Care</td>
<td>CA</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>College Health Enterprises</td>
<td>CA</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>HealthPlus</td>
<td>TX</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>Alameda Medical Center</td>
<td>CA</td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>Loma Linda University Health</td>
<td>CA</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>Palomar Pomerado Health</td>
<td>CA</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>Vibra Healthcare</td>
<td>PA</td>
<td>23</td>
<td>21</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>Cottage Health System</td>
<td>CA</td>
<td>27</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>Community Medical Centers</td>
<td>CA</td>
<td>31</td>
<td>31</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>John Muir Health</td>
<td>CA</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3.3: Hospital Systems Operating in California \(N=55\)
<table>
<thead>
<tr>
<th>HMO</th>
<th>Enrollment</th>
<th>Share</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blue Cross of California</td>
<td>3,913,413</td>
<td>0.2605</td>
</tr>
<tr>
<td>California Physicians’ Service</td>
<td>2,617,096</td>
<td>0.1742</td>
</tr>
<tr>
<td>Health Net of California, Inc.</td>
<td>2,123,679</td>
<td>0.1414</td>
</tr>
<tr>
<td>PacifiCare of California</td>
<td>1,283,343</td>
<td>0.0854</td>
</tr>
<tr>
<td>Local Initiative Health Authority For L.A. County</td>
<td>773,455</td>
<td>0.0515</td>
</tr>
<tr>
<td>Aetna Health of California, Inc.</td>
<td>459,827</td>
<td>0.0306</td>
</tr>
<tr>
<td>Inland Empire Health Plan</td>
<td>382,265</td>
<td>0.0255</td>
</tr>
<tr>
<td>Orange County Health Authority</td>
<td>351,056</td>
<td>0.0234</td>
</tr>
<tr>
<td>Molina Healthcare of California</td>
<td>321,744</td>
<td>0.0214</td>
</tr>
<tr>
<td>Heritage Provider Network, Inc.</td>
<td>321,234</td>
<td>0.0214</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>15,020,203</strong></td>
<td><strong>1.0000</strong></td>
</tr>
</tbody>
</table>

Table 3.4: Top 10 Managed Care Companies
Figure 3.1: California Managed Care Organizations Ordered by Market Share.
<table>
<thead>
<tr>
<th>Category</th>
<th>Characteristic</th>
<th>Mean</th>
<th>S.D.</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insurer</td>
<td>Private</td>
<td>.635</td>
<td>.232</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Medicare</td>
<td>.365</td>
<td>.232</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Type</td>
<td>Managed Care</td>
<td>.620</td>
<td>.236</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Fee For Service</td>
<td>.365</td>
<td>.232</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Diagnosis</td>
<td>Length of stay (days)</td>
<td>4.151</td>
<td>6.0259</td>
<td>0</td>
<td>378</td>
</tr>
<tr>
<td>Diagnostic</td>
<td>Nervous System Diagnosis</td>
<td>.060</td>
<td>.057</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Category (Shares)</td>
<td>Ear, Nose, Mouth, &amp; Throat</td>
<td>.013</td>
<td>.013</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Respiratory System</td>
<td>.087</td>
<td>.079</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Circulatory System</td>
<td>.159</td>
<td>.134</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Digestive System</td>
<td>.108</td>
<td>.097</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Hepatobiliary System</td>
<td>.041</td>
<td>.040</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Musculoskeletal System</td>
<td>.100</td>
<td>.090</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Skin, Subcutaneous Tissue &amp; Breast</td>
<td>.028</td>
<td>.028</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Endocrine, Nutritional, &amp; Metabolic</td>
<td>.039</td>
<td>.038</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Kidney &amp; Urinary Tract</td>
<td>.041</td>
<td>.039</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Pregnancy &amp; Childbirth</td>
<td>.190</td>
<td>.153</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Blood &amp; Immunological</td>
<td>.012</td>
<td>.012</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Myeloproliferative</td>
<td>.013</td>
<td>.013</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Infectious &amp; Parasitic</td>
<td>.033</td>
<td>.031</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Alcohol-Drug</td>
<td>.005</td>
<td>.005</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Patient</td>
<td>Male</td>
<td>.319</td>
<td>.269</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Age &lt; 18</td>
<td>.058</td>
<td>.055</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>18 ≤ Age &lt; 35</td>
<td>.197</td>
<td>.158</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>35 ≤ Age &lt; 65</td>
<td>.467</td>
<td>.249</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>65 ≤ Age &lt; 75</td>
<td>.278</td>
<td>.922</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>White</td>
<td>.579</td>
<td>.243</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Black</td>
<td>.044</td>
<td>.042</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Asian</td>
<td>.070</td>
<td>.065</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Choice Set Size</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Travel Time (min. to chosen hospital)</td>
<td>20.66</td>
<td>15.05</td>
<td>1</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>Travel Time (min. to all hospitals)</td>
<td>35.69</td>
<td>13.19</td>
<td>1</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>Rural</td>
<td>.029</td>
<td>.028</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Income ($1,000s)</td>
<td>51.50</td>
<td>19.86</td>
<td>0</td>
<td>200</td>
</tr>
</tbody>
</table>

Table 3.5: Discharges \( (N = 2,027,323) \)
<table>
<thead>
<tr>
<th>Service Name</th>
<th>Diagnostic Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emergency Room</td>
<td>Unrestricted*</td>
</tr>
<tr>
<td>Positron emission tomography</td>
<td>Unrestricted</td>
</tr>
<tr>
<td>Magnetic resonance imaging</td>
<td>Unrestricted</td>
</tr>
<tr>
<td>Neurological services</td>
<td>Nervous System</td>
</tr>
<tr>
<td>Adult diagnostic/invasive catheterization</td>
<td>Circulatory System</td>
</tr>
<tr>
<td>Adult cardiac electrophysiology</td>
<td>Circulatory System</td>
</tr>
<tr>
<td>Esophageal impedance study</td>
<td>Digestive System</td>
</tr>
<tr>
<td>Bariatric/weight control services</td>
<td>Digestive System</td>
</tr>
<tr>
<td>Oncology services</td>
<td>Hepatobiliary System</td>
</tr>
<tr>
<td>Medical/surgical intensive care</td>
<td>Unrestricted</td>
</tr>
<tr>
<td>Orthopedic services</td>
<td>Musculoskeletal System</td>
</tr>
<tr>
<td>Burn care</td>
<td>Skin, Subcut. Tissue &amp; Breast</td>
</tr>
<tr>
<td>Nutrition programs</td>
<td>Endocrine, Nutrit., &amp; Metabolic</td>
</tr>
<tr>
<td>Obstetrics care</td>
<td>Pregnancy &amp; Childbirth</td>
</tr>
<tr>
<td>Ultrasound</td>
<td>Pregnancy &amp; Childbirth</td>
</tr>
<tr>
<td>Birthing Room</td>
<td>Pregnancy &amp; Childbirth</td>
</tr>
<tr>
<td>Chemotherapy</td>
<td>Unrestricted</td>
</tr>
<tr>
<td>Alcohol/drug abuse or dependency inpatient care</td>
<td>Alcohol-Drug</td>
</tr>
</tbody>
</table>

Table 3.6: Included Services by Diagnostic Category
<table>
<thead>
<tr>
<th>Parameter</th>
<th>2007</th>
<th>2008</th>
</tr>
</thead>
<tbody>
<tr>
<td>Travel Time</td>
<td>-0.1526 (0.0016)</td>
<td>-0.1684 (0.0013)</td>
</tr>
<tr>
<td>Teaching Hospital</td>
<td>0.3708 (0.0134)</td>
<td>0.2023 (0.0135)</td>
</tr>
<tr>
<td>Teaching Hospital \times Travel Time</td>
<td>0.0087 (0.0005)</td>
<td>0.0091 (0.0005)</td>
</tr>
<tr>
<td>Nurse-to-Bed</td>
<td>-0.3895 (0.0071)</td>
<td>0.0470 (0.0070)</td>
</tr>
<tr>
<td>Nurse-to-Bed \times Travel Time</td>
<td>0.0243 (0.0003)</td>
<td>0.0193 (0.0003)</td>
</tr>
<tr>
<td>Income \times Travel Time</td>
<td>0.0798 (9.46E-6)</td>
<td>0.0111 (9.64E-6)</td>
</tr>
<tr>
<td>Length of Stay \times Travel Time</td>
<td>-0.0014 (1.81E-4)</td>
<td>0.0004 (1.87E-5)</td>
</tr>
<tr>
<td>For-Profit</td>
<td>0.0881 (0.0223)</td>
<td>0.1911 (0.0194)</td>
</tr>
<tr>
<td>For-Profit \times Travel Time</td>
<td>-0.0199 (0.0017)</td>
<td>-0.0322 (0.0009)</td>
</tr>
<tr>
<td>Government (State, County)</td>
<td>-0.7222 (0.0207)</td>
<td>-0.6214 (0.0205)</td>
</tr>
<tr>
<td>Government (State, County) \times Travel Time</td>
<td>0.0044 (0.0008)</td>
<td>-0.0023 (0.0008)</td>
</tr>
<tr>
<td>Rural \times Travel Time</td>
<td>0.0091 (0.0005)</td>
<td>-0.0589 (0.0004)</td>
</tr>
<tr>
<td>18 \leq \text{Age} &lt; 35 \times Travel Time</td>
<td>-0.0057 (0.0013)</td>
<td>-0.0109 (0.0013)</td>
</tr>
<tr>
<td>35 \leq \text{Age} &lt; 65 \times Travel Time</td>
<td>-0.0185 (0.0011)</td>
<td>-0.0246 (0.0011)</td>
</tr>
<tr>
<td>65 \leq \text{Age} &lt; 75 \times Travel Time</td>
<td>-0.0249 (0.0011)</td>
<td>-0.0333 (0.0011)</td>
</tr>
<tr>
<td>Female \times Travel Time</td>
<td>-0.0089 (0.0004)</td>
<td>-0.0077 (0.0004)</td>
</tr>
<tr>
<td>E.R. \times Travel Time</td>
<td>-0.0071 (0.0002)</td>
<td>-0.0031 (0.0002)</td>
</tr>
</tbody>
</table>

Table 3.7: Demand (#Chosen=2,027,323; #Not-Chosen=112,662,544)

Notes: All estimates are significant at the $p < .01$ level. Reported estimates represent the hospital and patient characteristics. All patient characteristics are interacted with time, which is in minutes. Omitted estimates include all estimates for the MDCs interacted with travel time (15 parameters), and all of the MDCs interacted with the hospital’s system status and travel time (45 parameters). All of the services listed in Table 3.6 are included individually and interacted with travel time in the demand estimate, but are omitted from the table.
<table>
<thead>
<tr>
<th>Major Diagnostic Cat.</th>
<th>Travel Time (min.)</th>
<th>Nor-For-Profit</th>
<th>For-Profit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Local</td>
<td>National</td>
</tr>
<tr>
<td>Nervous System</td>
<td>16</td>
<td>-0.0922</td>
<td>0.1198</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0122)</td>
<td>(0.0232)</td>
</tr>
<tr>
<td>Ear, Nose, Mouth</td>
<td>17</td>
<td>-0.0733</td>
<td>0.2338</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0256)</td>
<td>(0.0555)</td>
</tr>
<tr>
<td>Respiratory System</td>
<td>16</td>
<td>-0.1946</td>
<td>0.2586</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0100)</td>
<td>(0.0178)</td>
</tr>
<tr>
<td>Circulatory System</td>
<td>16</td>
<td>-0.0988</td>
<td>0.1405</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0087)</td>
<td>(0.0145)</td>
</tr>
<tr>
<td>Digestive System</td>
<td>16</td>
<td>-0.1333</td>
<td>0.3586</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0107)</td>
<td>(0.0185)</td>
</tr>
<tr>
<td>Hepatobiliary System</td>
<td>16</td>
<td>-0.0314</td>
<td>0.3854</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0166)</td>
<td>(0.0292)</td>
</tr>
<tr>
<td>Musculoskeletal System</td>
<td>18</td>
<td>0.0269</td>
<td>0.5451</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0094)</td>
<td>(0.0169)</td>
</tr>
<tr>
<td>Skin, Subcutaneous Tissue</td>
<td>16</td>
<td>-0.1417</td>
<td>0.2912</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0163)</td>
<td>(0.0338)</td>
</tr>
<tr>
<td>Endocrine, Nutritional</td>
<td>17</td>
<td>-0.0712</td>
<td>0.2479</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0155)</td>
<td>(0.0306)</td>
</tr>
<tr>
<td>Kidney</td>
<td>16</td>
<td>0.0518</td>
<td>0.3423</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0118)</td>
<td>(0.0262)</td>
</tr>
<tr>
<td>Pregnancy &amp; Childbirth</td>
<td>16</td>
<td>0.0137</td>
<td>0.4275</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0200)</td>
<td>(0.0397)</td>
</tr>
<tr>
<td>Blood &amp; Immunologic</td>
<td>17</td>
<td>0.0375</td>
<td>0.3771</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0220)</td>
<td>(0.0442)</td>
</tr>
<tr>
<td>Myeloproliferative</td>
<td>22</td>
<td>-0.1675</td>
<td>0.6042</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0237)</td>
<td>(0.0399)</td>
</tr>
<tr>
<td>Infectious &amp; Parasitic</td>
<td>15</td>
<td>-0.0638</td>
<td>0.2428</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0134)</td>
<td>(0.0231)</td>
</tr>
<tr>
<td>Alcohol-Drug</td>
<td>15</td>
<td>-0.1852</td>
<td>-0.5160</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0396)</td>
<td>(0.0616)</td>
</tr>
</tbody>
</table>

Table 3.8: Incremental Patient Value (\textit{Utils}) for Hospital System Membership

Notes: Standard errors are reported in parenthesis. All values are derived from the 2008 demand estimation. Travel Time represents the median travel time for each MDC. Each point estimate represents the difference in value for a hospital between being in a system for a patient with the median travel time. System status is categorized as being local (in-state only system) and national (having hospitals out of state).
Table 3.9: Cost Estimates: Variable Cost Characteristics

Notes: LOS is the total number in-patient days for either Medicare or the privately insured. Omitted coefficients include estimates for all other in-patient days and government interactions with outputs. All outputs are in logs as well as the number of staff, the number of beds, and the dollar amount of fixed assets. All specifications include time and HSA fixed effects and have an adjusted $R^2$ greater than 0.97.

Significance Levels: ***$p < .01$, **$p < .05$, *$p < .1$
<table>
<thead>
<tr>
<th></th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>C5</th>
</tr>
</thead>
<tbody>
<tr>
<td>For-Profit</td>
<td>-3.8590</td>
<td>-4.1208</td>
<td>-4.0434</td>
<td>-5.9317</td>
<td>-5.6647</td>
</tr>
<tr>
<td></td>
<td>(0.8595)</td>
<td>(0.8512)</td>
<td>(0.8353)</td>
<td>(0.9629)</td>
<td>(0.9865)</td>
</tr>
<tr>
<td>Rural</td>
<td>-0.1287</td>
<td>-0.1298</td>
<td>-0.1091</td>
<td>-0.1385</td>
<td>-0.1246</td>
</tr>
<tr>
<td></td>
<td>(0.0170)</td>
<td>(0.0168)</td>
<td>(0.0168)</td>
<td>(0.0168)</td>
<td>(0.0170)</td>
</tr>
<tr>
<td>Teaching</td>
<td>0.1195</td>
<td>0.0887</td>
<td>-0.3070</td>
<td>0.0663</td>
<td>-0.0943</td>
</tr>
<tr>
<td></td>
<td>(0.0244)</td>
<td>(0.0246)</td>
<td>(4.1142)</td>
<td>(0.0248)</td>
<td>(4.1027)</td>
</tr>
<tr>
<td>System</td>
<td>0.0723</td>
<td>0.0554</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0105)</td>
<td>(0.0106)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PET</td>
<td>-0.0014</td>
<td>-0.0049</td>
<td>-0.0144</td>
<td>-0.0014</td>
<td>-0.0087</td>
</tr>
<tr>
<td></td>
<td>(0.0118)</td>
<td>(0.0117)</td>
<td>(0.0115)</td>
<td>(0.0114)</td>
<td>(0.0114)</td>
</tr>
<tr>
<td>MRI</td>
<td>-0.0109</td>
<td>-0.0103</td>
<td>0.0013</td>
<td>-0.0074</td>
<td>0.0019</td>
</tr>
<tr>
<td></td>
<td>(0.0116)</td>
<td>(0.0115)</td>
<td>(0.0113)</td>
<td>(0.0113)</td>
<td>(0.0112)</td>
</tr>
<tr>
<td>%Medicare</td>
<td>-0.3012</td>
<td>-0.2661</td>
<td>-0.2607</td>
<td>-0.2237</td>
<td>-0.2263</td>
</tr>
<tr>
<td></td>
<td>(0.0509)</td>
<td>(0.0507)</td>
<td>(0.0506)</td>
<td>(0.0510)</td>
<td>(0.0515)</td>
</tr>
<tr>
<td>%E.R.</td>
<td>-0.2040</td>
<td>-0.1914</td>
<td>-0.1715</td>
<td>-0.1503</td>
<td>-0.1519</td>
</tr>
<tr>
<td></td>
<td>(0.0411)</td>
<td>(0.0407)</td>
<td>(0.0404)</td>
<td>(0.0407)</td>
<td>(0.0408)</td>
</tr>
<tr>
<td>Physicians’ Grp.</td>
<td>0.0241</td>
<td>0.0573*</td>
<td>0.0464</td>
<td>0.0297</td>
<td>0.0289</td>
</tr>
<tr>
<td></td>
<td>(0.0280)</td>
<td>(0.0281)</td>
<td>(0.0274)</td>
<td>(0.0286)</td>
<td>(0.0282)</td>
</tr>
<tr>
<td># Staff</td>
<td>-0.6919</td>
<td>-0.6555</td>
<td>-0.7004</td>
<td>-0.5305</td>
<td>-0.5651</td>
</tr>
<tr>
<td></td>
<td>(0.0998)</td>
<td>(0.0989)</td>
<td>(0.1037)</td>
<td>(0.1021)</td>
<td>(0.1086)</td>
</tr>
<tr>
<td># Staff Sqrd.</td>
<td>0.0700</td>
<td>0.0696</td>
<td>0.0720</td>
<td>0.0594</td>
<td>0.0614</td>
</tr>
<tr>
<td></td>
<td>(0.0082)</td>
<td>(0.0081)</td>
<td>(0.0085)</td>
<td>(0.0083)</td>
<td>(0.0089)</td>
</tr>
<tr>
<td>Fixed Assets ($)</td>
<td>-0.2384</td>
<td>-0.2609</td>
<td>-0.2887</td>
<td>-0.2730</td>
<td>-0.3077</td>
</tr>
<tr>
<td></td>
<td>(0.0520)</td>
<td>(0.0516)</td>
<td>(0.0508)</td>
<td>(0.0518)</td>
<td>(0.0518)</td>
</tr>
<tr>
<td>Fixed Assets Sqrd.</td>
<td>$</td>
<td>0.0102*</td>
<td>0.0110*</td>
<td>0.0118*</td>
<td>0.0112*</td>
</tr>
<tr>
<td>($^2$)</td>
<td>(0.0016)</td>
<td>(0.0016)</td>
<td>(0.0016)</td>
<td>(0.0016)</td>
<td>(0.0016)</td>
</tr>
<tr>
<td># Beds</td>
<td>0.2173</td>
<td>0.2020</td>
<td>0.1966</td>
<td>0.1609</td>
<td>0.1655</td>
</tr>
<tr>
<td></td>
<td>(0.0409)</td>
<td>(0.0405)</td>
<td>(0.0398)</td>
<td>(0.0404)</td>
<td>(0.0402)</td>
</tr>
<tr>
<td># Beds Sqrd.</td>
<td>-0.0234</td>
<td>-0.0228*</td>
<td>-0.0224</td>
<td>-0.0188</td>
<td>-0.0194</td>
</tr>
<tr>
<td></td>
<td>(0.0047)</td>
<td>(0.0047)</td>
<td>(0.0046)</td>
<td>(0.0047)</td>
<td>(0.0047)</td>
</tr>
<tr>
<td></td>
<td>(0.5401)</td>
<td>(0.5358)</td>
<td>(0.5396)</td>
<td>(0.5783)</td>
<td>(0.5937)</td>
</tr>
</tbody>
</table>

Table 3.10: Cost Estimates: Fixed Cost Characteristics

Notes: Estimates are a continuation from Table 3.9. Significance Levels:
***p < .01, ** p < .05, * p < .1
Table 3.11: Estimated Cost Differential for Not-For-Profit and For-Profit Hospitals

<table>
<thead>
<tr>
<th>Cost Model</th>
<th>Total Difference</th>
<th>Diff. Due to Characteristics</th>
<th>Diff. Due to Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>285.4%</td>
<td>268.5%</td>
<td>16.9%</td>
</tr>
<tr>
<td>C2</td>
<td>321.2%</td>
<td>304.4%</td>
<td>16.7%</td>
</tr>
<tr>
<td>C3</td>
<td>274.7%</td>
<td>259.3%</td>
<td>15.4%</td>
</tr>
<tr>
<td>C4</td>
<td>291.0%</td>
<td>275.6%</td>
<td>15.4%</td>
</tr>
<tr>
<td>C5</td>
<td>253.5%</td>
<td>234.6%</td>
<td>18.9%</td>
</tr>
</tbody>
</table>

Total Difference $= \ln \hat{C}^{NP} - \ln \hat{C}^{FP}$

Difference due to characteristics $= \sum [X_{i}^{NP} - X_{i}^{FP}] \hat{\beta}_i$

Difference due to parameters $= \sum X_{i}^{NP} \hat{\beta}_i$

Predictions are evaluated at the sample means for not-for-profit hospitals, $X^{SYS}$, and for-profit hospitals $X^N$. The $\hat{\beta}_i$ are the coefficient estimates reported in Table 3.9.
<table>
<thead>
<tr>
<th>Cost Model</th>
<th>Total Difference</th>
<th>Diff. Due to Characteristics</th>
<th>Diff. Due to Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>C2</td>
<td>40.7%</td>
<td>33.5%</td>
<td>7.2%</td>
</tr>
<tr>
<td>C3</td>
<td>27.7%</td>
<td>22.2%</td>
<td>5.5%</td>
</tr>
<tr>
<td>C4</td>
<td>32.7%</td>
<td>29.3%</td>
<td>3.4%</td>
</tr>
<tr>
<td>C5</td>
<td>26.7%</td>
<td>24.3%</td>
<td>2.3%</td>
</tr>
</tbody>
</table>

Table 3.12: Estimated Cost Differential for System and Non-System Hospitals

\[
\text{Total Difference} = \ln \hat{C}^{SYS} - \ln \hat{C}^{N}
\]
\[
\text{Difference due to characteristics} = \sum [X_{i}^{SYS} - X_{i}^{N}] \hat{\beta}_i
\]
\[
\text{Difference due to parameters} = \sum X_{i}^{SYS} \hat{\beta}_i
\]

Predictions are evaluated at the sample means for system hospitals, \(X^{SYS}\), and non-system hospitals \(X^{N}\). The \(\hat{\beta}_i\) are the coefficient estimates reported in Table 3.9.
<table>
<thead>
<tr>
<th>Dependent Var. = $\Delta \Pi_h$</th>
<th>Cost Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C1</td>
</tr>
<tr>
<td>Base Bargaining Pwr.</td>
<td>0.3101</td>
</tr>
<tr>
<td>(0.4439)</td>
<td>(0.1607)</td>
</tr>
<tr>
<td>CPHO</td>
<td>0.9274*</td>
</tr>
<tr>
<td>(0.5015)</td>
<td>(0.2027)</td>
</tr>
<tr>
<td>OPHO</td>
<td>-0.2462</td>
</tr>
<tr>
<td>(0.2660)</td>
<td>(0.6560)</td>
</tr>
<tr>
<td>IPA</td>
<td>0.0542</td>
</tr>
<tr>
<td>(0.1102)</td>
<td>(0.1161)</td>
</tr>
<tr>
<td>GPWW</td>
<td>0.3639**</td>
</tr>
<tr>
<td>(0.1579)</td>
<td>(0.1563)</td>
</tr>
<tr>
<td>Teaching</td>
<td>0.3119</td>
</tr>
<tr>
<td>(0.2517)</td>
<td>(0.6857)</td>
</tr>
<tr>
<td>Rural</td>
<td>-0.3236**</td>
</tr>
<tr>
<td>(0.1671)</td>
<td>(0.1338)</td>
</tr>
<tr>
<td># Beds (/100)</td>
<td>-0.0583</td>
</tr>
<tr>
<td>(0.2670)</td>
<td>(0.0754)</td>
</tr>
<tr>
<td># Beds Sqrd.</td>
<td>0.0170</td>
</tr>
<tr>
<td>(0.0461)</td>
<td>(0.0130)</td>
</tr>
<tr>
<td>For Profit</td>
<td>-0.1076</td>
</tr>
<tr>
<td>(0.0358)</td>
<td>(0.0979)</td>
</tr>
<tr>
<td>Market Share</td>
<td>1.6050***</td>
</tr>
<tr>
<td>(0.5495)</td>
<td>(0.5232)</td>
</tr>
<tr>
<td>Market Share Sqrd.</td>
<td>-1.8462***</td>
</tr>
<tr>
<td>(0.5248)</td>
<td>(0.5074)</td>
</tr>
<tr>
<td>System Mem.</td>
<td>0.1365</td>
</tr>
<tr>
<td>(0.1154)</td>
<td>(0.0918)</td>
</tr>
<tr>
<td>BCBS Mrkt. Share</td>
<td>-0.5202*</td>
</tr>
<tr>
<td>(0.9000)</td>
<td>(0.3480)</td>
</tr>
</tbody>
</table>

$\gamma_p^{-1}$

| Adj. R² | 0.7891  | 0.7682  | 0.7493  | 0.7373  | 0.7483  | 0.7234  | 0.7269  | 0.7060  | 0.7303  | 0.6983  |

161

Table 3.13: Determinants of Bargaining Power
Dependent Var. = ΔΠₜₙ

Regression Specification

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant (Base Bargaining Pwr.)</td>
<td>0.5448**</td>
<td>0.5346**</td>
<td>0.5678***</td>
<td>0.4659**</td>
<td>0.4958**</td>
<td>0.4774**</td>
<td>0.5160***</td>
</tr>
<tr>
<td></td>
<td>(0.2214)</td>
<td>(0.2224)</td>
<td>(0.2161)</td>
<td>(0.2122)</td>
<td>(0.2220)</td>
<td>(0.1940)</td>
<td>(0.1976)</td>
</tr>
<tr>
<td>Market Share (≤ 10mi.)</td>
<td>1.2111**</td>
<td>1.1899**</td>
<td>1.2234**</td>
<td>1.2952***</td>
<td>1.2520**</td>
<td>1.4979***</td>
<td>1.4673***</td>
</tr>
<tr>
<td></td>
<td>(0.5339)</td>
<td>(0.5329)</td>
<td>(0.5537)</td>
<td>(0.5000)</td>
<td>(0.5362)</td>
<td>(0.5569)</td>
<td>(0.5224)</td>
</tr>
<tr>
<td>Market Share Sqrd. (≤ 10mi.)</td>
<td>-1.5502***</td>
<td>-1.5212***</td>
<td>-1.6162***</td>
<td>-1.5581***</td>
<td>-1.5473***</td>
<td>-1.7387***</td>
<td>-1.6815***</td>
</tr>
<tr>
<td></td>
<td>(0.5147)</td>
<td>(0.5201)</td>
<td>(0.5204)</td>
<td>(0.4572)</td>
<td>(0.4825)</td>
<td>(0.5127)</td>
<td>(0.4617)</td>
</tr>
<tr>
<td>FP×System Member</td>
<td>0.2481*</td>
<td>0.2494*</td>
<td>0.0885</td>
<td>0.0838</td>
<td>-0.1435</td>
<td>-0.6301**</td>
<td>-0.6320***</td>
</tr>
<tr>
<td></td>
<td>(0.1462)</td>
<td>(0.1469)</td>
<td>(0.2449)</td>
<td>(0.1751)</td>
<td>(0.2535)</td>
<td>(0.2428)</td>
<td>(0.2407)</td>
</tr>
<tr>
<td>NFP×System Member</td>
<td>0.1237</td>
<td>0.1174</td>
<td>0.2697**</td>
<td>0.0508</td>
<td>0.1507</td>
<td>0.1854</td>
<td>0.1837</td>
</tr>
<tr>
<td></td>
<td>(0.1068)</td>
<td>(0.1158)</td>
<td>(0.1172)</td>
<td>(0.1078)</td>
<td>(0.2020)</td>
<td>(0.1242)</td>
<td>(0.1293)</td>
</tr>
<tr>
<td>Includes Teaching Hosp.</td>
<td>-0.0201</td>
<td>-0.0389</td>
<td>-0.0314</td>
<td>0.0127</td>
<td>0.0031</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0850)</td>
<td>(0.0889)</td>
<td>(0.0868)</td>
<td>(0.0887)</td>
<td>(0.1872)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FP×Multi-market System</td>
<td>0.2046</td>
<td>0.2620</td>
<td>0.2137**</td>
<td>0.0508</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.2529)</td>
<td>(0.2287)</td>
<td>(0.2336)</td>
<td>(0.2338)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NFP×Multi-market System</td>
<td>-0.2129**</td>
<td>-0.1209</td>
<td>-0.1044</td>
<td>-0.0902</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0855)</td>
<td>(0.1266)</td>
<td>(0.0898)</td>
<td>(0.0968)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FP×Sys. Market Share (≤ 10mi.)</td>
<td>2.3377*</td>
<td>2.4422*</td>
<td>2.6239**</td>
<td>2.5880**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.1891)</td>
<td>(1.2770)</td>
<td>(1.1565)</td>
<td>(1.2022)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NFP×Sys. Market Share (≤ 10mi.)</td>
<td>0.8361**</td>
<td>0.6238</td>
<td>0.2955</td>
<td>0.3149</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.3409)</td>
<td>(0.7532)</td>
<td>(0.3516)</td>
<td>(0.3323)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FP×Close Partner (≤ 2.5mi.)</td>
<td>1.0362*</td>
<td>1.0595**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.4131)</td>
<td>(0.4044)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NFP×Close Partner (≤ 2.5mi.)</td>
<td>0.2548*</td>
<td>0.2536*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.1311)</td>
<td>(0.1305)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Large System Controls</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Adj. R²</td>
<td>0.8034</td>
<td>0.8034</td>
<td>0.8095</td>
<td>0.8135</td>
<td>0.8159</td>
<td>0.8251</td>
<td>0.8286</td>
</tr>
</tbody>
</table>

Table 3.14: System Characteristics and Bargaining Power
<table>
<thead>
<tr>
<th>Dependent Var. = $\Delta \Pi_h$</th>
<th>Regression Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
</tr>
<tr>
<td>Constant (Base Bargaining Pwr.)</td>
<td>0.3697*</td>
</tr>
<tr>
<td></td>
<td>(0.1942)</td>
</tr>
<tr>
<td>Market Share ($\leq$ 10mi.)</td>
<td>0.8924*</td>
</tr>
<tr>
<td></td>
<td>(0.4993)</td>
</tr>
<tr>
<td>Market Share Sqrd. ($\leq$ 10mi.)</td>
<td>-1.1920**</td>
</tr>
<tr>
<td></td>
<td>(0.4934)</td>
</tr>
<tr>
<td>FP $\times$ System Member</td>
<td>0.1115</td>
</tr>
<tr>
<td></td>
<td>(0.1745)</td>
</tr>
<tr>
<td>NFP $\times$ System Member</td>
<td>0.0476</td>
</tr>
<tr>
<td></td>
<td>(0.0869)</td>
</tr>
<tr>
<td>Includes Teaching Hosp.</td>
<td>0.0576</td>
</tr>
<tr>
<td></td>
<td>(0.0775)</td>
</tr>
<tr>
<td>FP $\times$ Multi-market System</td>
<td>0.1692</td>
</tr>
<tr>
<td></td>
<td>(0.1953)</td>
</tr>
<tr>
<td>NFP $\times$ Multi-market System</td>
<td>-0.0989</td>
</tr>
<tr>
<td></td>
<td>(0.0652)</td>
</tr>
<tr>
<td>FP $\times$ Sys. Market Share ($\leq$ 10mi.)</td>
<td>1.8853**</td>
</tr>
<tr>
<td></td>
<td>(0.8898)</td>
</tr>
<tr>
<td>NFP $\times$ Sys. Market Share ($\leq$ 10mi.)</td>
<td>0.3147</td>
</tr>
<tr>
<td></td>
<td>(0.2601)</td>
</tr>
<tr>
<td>FP $\times$ Close Partner ($\leq$ 2.5mi.)</td>
<td>0.7715**</td>
</tr>
<tr>
<td></td>
<td>(0.3485)</td>
</tr>
<tr>
<td>NFP $\times$ Close Partner ($\leq$ 2.5mi.)</td>
<td>0.1654</td>
</tr>
<tr>
<td></td>
<td>(0.1203)</td>
</tr>
<tr>
<td>Large System Controls</td>
<td>No</td>
</tr>
<tr>
<td>Adj. $R^2$</td>
<td>0.7809</td>
</tr>
</tbody>
</table>

Table 3.15: System Characteristics and Bargaining Power Adjusted for System Market Power
<table>
<thead>
<tr>
<th>Variable</th>
<th>50%</th>
<th>25%</th>
<th>15%</th>
<th>10%</th>
<th>1%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hosp. Pwr.</td>
<td>0.5300**</td>
<td>0.3609*</td>
<td>0.5290***</td>
<td>0.3613**</td>
<td>0.5485***</td>
</tr>
<tr>
<td>(0.2289)</td>
<td>(0.1910)</td>
<td>(0.2129)</td>
<td>(0.1826)</td>
<td>(0.2133)</td>
<td>(0.1815)</td>
</tr>
<tr>
<td>CPHO</td>
<td>0.1553</td>
<td>0.0890</td>
<td>0.1349</td>
<td>0.0903</td>
<td>0.1292</td>
</tr>
<tr>
<td>(0.3459)</td>
<td>(0.2882)</td>
<td>(0.3230)</td>
<td>(0.2712)</td>
<td>(0.3502)</td>
<td>(0.2654)</td>
</tr>
<tr>
<td>OPHO</td>
<td>-0.1770</td>
<td>-0.1303</td>
<td>-0.1732</td>
<td>-0.1288</td>
<td>-0.1721</td>
</tr>
<tr>
<td>(0.2210)</td>
<td>(0.2366)</td>
<td>(0.2140)</td>
<td>(0.2266)</td>
<td>(0.2118)</td>
<td>(0.2319)</td>
</tr>
<tr>
<td>IPA</td>
<td>-0.0030</td>
<td>-0.0110</td>
<td>-0.0015</td>
<td>-0.0091</td>
<td>-0.0010</td>
</tr>
<tr>
<td>(0.0913)</td>
<td>(0.0787)</td>
<td>(0.0882)</td>
<td>(0.0748)</td>
<td>(0.0872)</td>
<td>(0.0740)</td>
</tr>
<tr>
<td>GPWW</td>
<td>0.3451***</td>
<td>0.2776***</td>
<td>0.3286***</td>
<td>0.2570**</td>
<td>0.3214***</td>
</tr>
<tr>
<td>(0.1160)</td>
<td>(0.1368)</td>
<td>(0.1108)</td>
<td>(0.1311)</td>
<td>(0.1067)</td>
<td>(0.1252)</td>
</tr>
<tr>
<td>Teaching</td>
<td>0.2388</td>
<td>0.2323</td>
<td>0.2343</td>
<td>0.2273</td>
<td>0.2329</td>
</tr>
<tr>
<td>(0.2042)</td>
<td>(0.2318)</td>
<td>(0.1992)</td>
<td>(0.2268)</td>
<td>(0.1991)</td>
<td>(0.2355)</td>
</tr>
<tr>
<td>Rural</td>
<td>-0.2726**</td>
<td>-0.1404**</td>
<td>-0.2621**</td>
<td>-0.1355**</td>
<td>-0.2587**</td>
</tr>
<tr>
<td>(0.1138)</td>
<td>(0.0908)</td>
<td>(0.1104)</td>
<td>(0.0870)</td>
<td>(0.1105)</td>
<td>(0.0859)</td>
</tr>
<tr>
<td># Beds (/100)</td>
<td>-0.0995</td>
<td>-0.0208</td>
<td>-0.0949</td>
<td>-0.0203</td>
<td>-0.0934</td>
</tr>
<tr>
<td>(0.1199)</td>
<td>(0.0871)</td>
<td>(0.1106)</td>
<td>(0.0810)</td>
<td>(0.1105)</td>
<td>(0.0792)</td>
</tr>
<tr>
<td># Beds Sqrd. (/100²)</td>
<td>0.0217</td>
<td>0.0213</td>
<td>0.0209</td>
<td>0.0207</td>
<td>0.0206</td>
</tr>
<tr>
<td>(0.0211)</td>
<td>(0.0157)</td>
<td>(0.0194)</td>
<td>(0.0146)</td>
<td>(0.0195)</td>
<td>(0.0143)</td>
</tr>
<tr>
<td>For Profit</td>
<td>-0.0044</td>
<td>-0.0511</td>
<td>-0.0064</td>
<td>-0.0071</td>
<td>-0.0071</td>
</tr>
<tr>
<td>(0.0927)</td>
<td>(0.0972)</td>
<td>(0.0973)</td>
<td>(0.0948)</td>
<td>(0.0973)</td>
<td>(0.0843)</td>
</tr>
<tr>
<td>Market Share</td>
<td>1.2954***</td>
<td>0.9653***</td>
<td>1.2506***</td>
<td>0.9216***</td>
<td>1.2352***</td>
</tr>
<tr>
<td>(0.5551)</td>
<td>(0.5288)</td>
<td>(0.5829)</td>
<td>(0.4938)</td>
<td>(0.5269)</td>
<td>(0.4877)</td>
</tr>
<tr>
<td>Market Share Sqrd.</td>
<td>-1.6324***</td>
<td>-1.2659***</td>
<td>-1.5854***</td>
<td>-1.2191***</td>
<td>-1.5693***</td>
</tr>
<tr>
<td>(5.539)</td>
<td>(0.5222)</td>
<td>(0.5151)</td>
<td>(0.4883)</td>
<td>(0.5096)</td>
<td>(0.4829)</td>
</tr>
<tr>
<td>System Mem.</td>
<td>0.1431</td>
<td>0.0538</td>
<td>0.1360</td>
<td>0.0532</td>
<td>0.1336</td>
</tr>
<tr>
<td>(0.1043)</td>
<td>(0.0868)</td>
<td>(0.1002)</td>
<td>(0.0816)</td>
<td>(0.0989)</td>
<td>(0.0805)</td>
</tr>
<tr>
<td>BCBS Mkrt. Share</td>
<td>-0.5143*</td>
<td>-0.3072*</td>
<td>-0.4970*</td>
<td>-0.2940</td>
<td>-0.4912*</td>
</tr>
<tr>
<td>(0.2832)</td>
<td>(0.2466)</td>
<td>(0.2749)</td>
<td>(0.2317)</td>
<td>(0.2766)</td>
<td>(0.2519)</td>
</tr>
</tbody>
</table>

| Adjusted R²                      | 0.7952 | 0.7706 | 0.8011 | 0.7782 | 0.8048 | 0.7809 | 0.8042 | 0.7821 | 0.7632 | 0.7635 |

Table 3.16: Determinants of Bargaining and Patient Size
<table>
<thead>
<tr>
<th>Model</th>
<th>75th percentile</th>
<th>50th Percentile</th>
<th>25th Percentile</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%</td>
<td>$</td>
<td>%</td>
<td>$</td>
</tr>
<tr>
<td>A</td>
<td>4.15</td>
<td>127.79</td>
<td>1.93</td>
<td>62.23</td>
</tr>
<tr>
<td>B</td>
<td>4.82</td>
<td>141.14</td>
<td>2.27</td>
<td>68.29</td>
</tr>
<tr>
<td>C</td>
<td>4.87</td>
<td>143.85</td>
<td>2.02</td>
<td>65.52</td>
</tr>
<tr>
<td>D</td>
<td>4.26</td>
<td>130.97</td>
<td>2.13</td>
<td>62.06</td>
</tr>
<tr>
<td>E</td>
<td>5.04</td>
<td>142.69</td>
<td>2.14</td>
<td>66.52</td>
</tr>
<tr>
<td>F</td>
<td>6.14</td>
<td>174.63</td>
<td>2.66</td>
<td>82.02</td>
</tr>
<tr>
<td>G</td>
<td>6.35</td>
<td>188.26</td>
<td>2.65</td>
<td>78.72</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>75th percentile</th>
<th>50th Percentile</th>
<th>25th Percentile</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%</td>
<td>$</td>
<td>%</td>
<td>$</td>
</tr>
<tr>
<td>A</td>
<td>19.62</td>
<td>436.74</td>
<td>12.48</td>
<td>338.27</td>
</tr>
<tr>
<td>B</td>
<td>26.46</td>
<td>618.81</td>
<td>17.80</td>
<td>428.44</td>
</tr>
<tr>
<td>C</td>
<td>24.77</td>
<td>605.46</td>
<td>14.55</td>
<td>378.87</td>
</tr>
<tr>
<td>D</td>
<td>17.46</td>
<td>485.94</td>
<td>9.55</td>
<td>268.51</td>
</tr>
<tr>
<td>E</td>
<td>22.70</td>
<td>559.89</td>
<td>12.23</td>
<td>325.07</td>
</tr>
<tr>
<td>F</td>
<td>34.77</td>
<td>825.93</td>
<td>19.54</td>
<td>508.90</td>
</tr>
<tr>
<td>G</td>
<td>32.13</td>
<td>798.10</td>
<td>14.89</td>
<td>417.03</td>
</tr>
</tbody>
</table>

Table 3.17: Markup of Average Daily Reimbursement

Notes: Markups represent the increase in a hospital’s average daily reimbursement rate that is attributable to system membership. The increase from Market Power represents the part of the markup that is attributable to the system adjusting an MCO’s threatpoint by threatening to withdraw all of the system hospitals from the MCO’s network. The increase from Bargaining Power represents the part of the markup that is attributable to the system extracting a larger share of the generated surplus conditional on the surplus increase due to Market Power. All markups are calculated under the condition that the entire system withdraws from a Network when there is a breakdown in bargaining.
Appendix A: Chapter 1 Proofs

PROOF OF PROPOSITION 1

We start by establishing the concavity of the firm’s profit functions. The first-order condition of the firm’s profit when higher types prefer the drug over their outside option is given as:

\[1 - F(\hat{\theta}(p)) - pf(\hat{\theta}(p))\hat{\theta}'(p) = 0. \tag{A-1}\]

Using the relationship \(1 - F(\hat{\theta}(p)) = pf(\hat{\theta}(p))\hat{\theta}'(p)\), the second-order condition can be expressed as:

\[
\frac{d^2\Pi}{dp^2} = \hat{\theta}'(p) \left[ -f(\hat{\theta}(p))^2 - f'(\hat{\theta}(p))(1 - F(\hat{\theta}(p))) \right] - f(\hat{\theta}(p))(p\hat{\theta}''(p) + \hat{\theta}'(p)). \tag{A-2}\]

The term in brackets is negative by the monotone hazard rate property and \(\hat{\theta}'(p) > 0\) since higher types choose the drug and lower types choose the outside option. Therefore we must determine \(\text{sign}[f(\hat{\theta}(p))(p\hat{\theta}''(p) + \hat{\theta}'(p))]\) to sign the second-order condition. Using implicit differentiation the derivatives \(\hat{\theta}'\) and \(\hat{\theta}''\) are given as:

\[
\hat{\theta}'(p) = \frac{\gamma}{1 - \gamma \overline{u}(\theta)} > 0, \tag{A-3}\]

\[
\hat{\theta}'(p) = \left[ \frac{\gamma}{1 - \gamma \overline{u}(\theta)} \right]^3 \overline{u}''(\theta) \geq 0, \tag{A-4}\]

where the signs are a result of the fact that higher types prefer the drug over their outside option and \(\overline{u}''(\theta) \geq 0\) by assumption. Therefore \(d^2\Pi/dp^2 < 0\) and the firm’s problem is concave. To complete the proof, there are two cases to consider as the demand for the drug differs depending on whether or not consumers are constrained or unconstrained.

When consumers are unconstrained the cut-off type is denoted as \(\hat{\theta}_S = p_S + \pi(\hat{\theta}_S)\). Recall also that \(\hat{\theta}_{NS} = \left[(1 + \mu(\Sigma - 1))/\mu \Sigma\right](p_{NS} + \pi(\hat{\theta}_{NS}))\). Using these definitions we can express the cut-off type generically as \(\hat{\theta}_s = \gamma_s(p_s + \pi(\hat{\theta}_s))\) where \(\gamma_S = 1\), \(\gamma_{NS} = \left[(1 + \mu(\Sigma - 1))/\mu \Sigma\right]\) and \(\gamma_S < \gamma_{NS}\) for all \(\mu \in (0,1)\) and for all \(\phi \in (0,1)\). To determine the relative size of \(p^*_S\) to \(p^*_NS\) we simply need to find \(\text{sign}[dp/d\gamma]\). Because of the concavity of the firm’s problem, from the Conjugate Pairs theorem, \(\text{sign}[dp/d\gamma] = \text{sign}[d^2\Pi_s/dpd\gamma]\). Let \(\hat{\theta}_s(p,\gamma)\).
solve \( \theta = \gamma(p + \overline{\mu}(\theta)) \) then \( d^2 \Pi_s / dp d\gamma \) is

\[
d^2 \Pi_s / dp d\gamma = -\hat{\theta}_p(p, \gamma)f(\hat{\theta}(p, \gamma)) - \hat{\theta}_\gamma(p, \gamma)p f(\hat{\theta}(p, \gamma)) - \hat{\theta}_\gamma(p, \gamma)p f(\hat{\theta}(p, \gamma))
\]

\[
= \{-\hat{\theta}_\gamma[(f(\cdot))^2 - (1 - F(\cdot))f'() - \hat{\theta}_\gamma p f(\cdot)^2] / f(),
\]

where the second line again follows from the first-order condition, \( (1 - F(\cdot)) = \hat{\theta}_p pf(\cdot) \). Therefore, we have

\[
\text{sign}[d^2 \Pi_s / dp d\gamma] = \text{sign}\left[\{-\hat{\theta}_\gamma[(f(\cdot))^2 - (1 - F(\cdot))f'()] - \hat{\theta}_\gamma p f(\cdot)^2\}\right]. \quad (A-5)
\]

By using implicit differentiation the signs for \( \hat{\theta}_p, \hat{\theta}_\gamma \) and \( \hat{\theta}_\gamma p \) can be shown to be equivalent, thus all are positive since \( \hat{\theta}_p \) is positive. The term in brackets in (A-5) is positive from the monotone hazard rate property of \( F \) and \( d^2 \Pi_s / dp d\gamma < 0 \) therefore \( dp / d\gamma < 0 \) and \( p^*_s > p^*_N S \).

To see the firm’s choice in market coverage, consider the firm’s choice variable to be that of choosing the profit-maximizing cut-off type. The firm’s optimization programs may now be expressed as

\[
\max_{\hat{\theta}_N S} (\alpha \hat{\theta}_N S - \overline{\mu}(\hat{\theta}_N S))(1 + \phi(T - 1))(1 - F(\hat{\theta}_N S)),
\]

\[
\max_{\hat{\theta}_S} (\hat{\theta}_S - \overline{\mu}(\hat{\theta}_S))\phi(T - 1)(1 - F(\hat{\theta}_S)).
\]

The first-order condition for the two profit functions can be expressed as

\[
(\gamma_s - \overline{\mu})(1 - F(\hat{\theta}_s)) - (\gamma_i \hat{\theta}_s - \overline{\mu}(\hat{\theta}_s)) f(\hat{\theta}_s) = 0, \quad (A-6)
\]

where \( s \in \{N S, S\}, \gamma_{N S} = \alpha \), and \( \gamma_S = 1 \). Using the fact that the firm’s profit functions are concave, from the Conjugate Pairs theorem we get, \( \text{sign}[d\theta / d\gamma] = \text{sign}[1 - F(\hat{\theta}_S) - \theta f(\hat{\theta}_s)] \).

From the FOC we have

\[
1 - F(\hat{\theta}_s) - \theta f(\hat{\theta}_s) = -\frac{\overline{\mu}(\hat{\theta}_s)}{\gamma_s}f(\hat{\theta}_s) < 0.
\]

Therefore \( d\theta / d\gamma > 0 \) and because \( \gamma_S > \gamma_{N S} \) it must be the case that \( \hat{\theta}_S^* < \hat{\theta}_{N S}^* \).

When consumers are constrained we need only look at how the firm’s choice of cut-off changes with \( \gamma \). Recall that when consumers are constrained, the cut-off type is denoted as

\[
\hat{\theta}_S = \left(\frac{\mu(\Sigma - 1)}{\mu \Sigma}\right) p_S + \left(\frac{1 + \mu(\Sigma - 1)}{\mu \Sigma}\right) \overline{\mu}(\hat{\theta}_S),
\]

\[
= \left(\frac{1 + \mu(\Sigma - 1)}{\mu \Sigma}\right) \left(\frac{\mu(\Sigma - 1)}{1 + \mu(\Sigma - 1)} p_S + \overline{\mu}(\hat{\theta}_S)\right), \quad (A-7)
\]

Let \( \alpha = 1/\gamma = \mu \Sigma/(1 + \mu(\Sigma - 1)) \) and \( \beta = (1 + \mu(\Sigma - 1))/\mu(\Sigma - 1) \). When consumers are constrained we can look at the proportion of the market that the firm optimally chooses to
cover. Rearranging (1.4) and (A-7) the firm’s optimization programs for the no-sampling and sampling cases can be expressed as

$$\max_{\theta_{NS}} (\alpha \hat{\theta}_{NS} - \overline{u}(\hat{\theta}_{NS})) (1 + \phi(T - 1)) (1 - F(\hat{\theta}_{NS})), $$

$$\max_{\theta_{NS}} \beta (\alpha \hat{\theta}_{S} - \overline{u}(\hat{\theta}_{S})) \phi(T - 1) (1 - F(\hat{\theta}_{S})).$$

It is clear from the first-order conditions that \( \hat{\theta}_{NS} = \hat{\theta}_{S} \) and the firm will cover the same proportion of the market with or without samples. Because \( \hat{\theta}_{NS} = \hat{\theta}_{S} \), it must be the case that \( p_{NS}^* = (1/\beta) p_{S}^* \) therefore \( p_{NS}^* < p_{S}^* \), completing the proof.

PROOF OF COROLLARY 1

When the firm cannot adjust the price paid by consumers then the market covered by the firm is determined by Eqs. (1.4) and (1.8). The proof for Proposition 1 shows that \( \hat{\theta}_{S} > 0 \), thus \( \hat{\theta}_{NS} > \hat{\theta}_{S} \) when consumers are unconstrained. When consumers are constrained, Eq. (A-7) shows that there is a coefficient \( \beta < 1 \) on \( p \) when the firm dispenses samples and it is clear that \( \partial \hat{\theta}/\partial \beta > 0 \). Because \( \beta = 1 \) when the firm does not provide samples we again have \( \hat{\theta}_{NS} > \hat{\theta}_{S} \) and the firm covers a larger proportion of the market when it dispenses samples.

PROOF OF PROPOSITION 2

At \( \phi = 0 \) we have \( E\Pi^S = -c_S \) and \( E\Pi^{NS} = 0 \) and at \( \phi = 1 \) we have \( E\Pi^{NS} = E\Pi^S + c_s + p_{NS}^*(1 - F(\hat{\theta}_{NS}(p_{NS}^*))) \), establishing that no-sampling is more profitable at either end-point and \( \Phi \subset [0, 1] \). In order to establish the convexity of \( \Phi \) we need to show that \( E\Pi^S(\phi) \) crosses \( E\Pi^{NS}(\phi) \) from below at most once. To demonstrate this we will first show that \( E\Pi^S(\phi) \) and \( E\Pi^{NS}(\phi) \) are weakly convex and then establish that this is a sufficient condition for a single-crossing.

Using the envelope theorem and similar substitutions as found in the proof for Proposition 1, \( d^2 E\Pi^{NS}/d\phi^2 \) can be expressed as

$$\frac{d^2 E\Pi^{NS}}{d\phi^2} = \frac{\partial^2 E\Pi^{NS}}{\partial \phi^2} + \frac{\partial^2 E\Pi^{NS}}{\partial \phi \partial p_{NS}^*} \frac{dp_{NS}^*}{d\phi}$$

$$= \{(T - 1)(1 - F(\hat{\theta}(p_{NS}^*))) - (1 + \phi(T - 1)) \frac{\partial \hat{\theta}}{\partial \phi} [f(\hat{\theta}(p_{NS}^*)) + \frac{\partial \hat{\theta}}{\partial p} p_{NS}^* f'(\hat{\theta}(p_{NS}^*))] \} \frac{dp_{NS}^*}{d\phi}$$

$$= \{(T - 1)(1 - F(\hat{\theta}(p_{NS}^*))) - (1 + \phi(T - 1)) \frac{\partial \hat{\theta}}{\partial \phi} [(f(\cdot))^2 + (1 - F(\cdot)) f'(\cdot)/f(\cdot)] \} \frac{dp_{NS}^*}{d\phi} > 0.$$

Therefore \( E\Pi^{NS} \) is strictly convex in \( \phi \). The derivation for \( d^2 E\Pi^S/d\phi^2 \) follows similarly and is weakly convex — when consumers are unconstrained \( \Pi^S \) is a linear function of \( \phi \) — and when consumers are constrained \( E\Pi^S \) is strictly convex.

To see that the property of convexity is sufficient to ensure that \( E\Pi^S \) crosses \( E\Pi^{NS} \) from below only once consider a case where \( E\Pi^S \) crosses \( E\Pi^{NS} \) from below. At the point of crossing \( dE\Pi^S/d\phi > dE\Pi^{NS}/d\phi \). Furthermore, because of the end-point conditions
we know that there exists another point at which $E\Pi^{NS}$ crosses $E\Pi^{S}$ from below and at this point $dE\Pi^{NS}/d\phi > dE\Pi^{S}/d\phi$, implying $d^2E\Pi^{NS}/d\phi^2 > d^2E\Pi^{S}/d\phi^2$. In order for $E\Pi^{S}$ to cross $E\Pi^{NS}$ from below again, given that the functions are continuous, we need $d^2E\Pi^{NS}/d\phi^2 < d^2E\Pi^{S}/d\phi^2$ for some interval, violating the convexity of the profit functions. Therefore, the sampling set $\Phi$ must be a convex set.

**PROOF OF LEMMA 1**

When the outside option of the consumers is increased the sampling and no-sampling profits are decreased; i.e., $dE\Pi^{s}/d\varpi < 0$ for $s \in \{S, NS\}$. Nevertheless, it is straightforward to show that

$$\frac{dE\Pi^{S}}{d\varpi} - \frac{dE\Pi^{NS}}{d\varpi} \geq \alpha Pf(\hat{\theta}^{NS}) \frac{d\hat{\theta}}{d\varpi} > 0,$$

and the difference between the sampling and no-sampling profits are increased increasing the size of the sampling set.

**PROOF OF LEMMA 2**

The proof for this statement follows directly from the proof for Proposition 4. The proof for Proposition 4 shows the firm’s sampling strategy and the consumers’ beliefs must be consistent in equilibrium. The only beliefs and sampling strategy which are consistent require a unique sampling cutoff, above which the firm always provides samples and below which it never provides samples. The unique cutoff is found by equating the firm’s expected profit from dispensing samples to the firm’s expected profit from not dispensing samples. The expected profits must take into account the consumers’ belief of the drug’s efficacy given that firm either dispenses or does not dispense samples. No cut-off efficacy exists when $E\Pi^{NS} > E\Pi^{S}$ for all $\phi \in [0, 1]$.

**PROOF OF LEMMA 3**

Let $G$ be the family of distributions having positive density everywhere on the supports $[0, 1]$. From Lemma 2 there exists a unique cutoff $\hat{\phi}$ such that for every $\phi > \hat{\phi}$ the firm will provide samples. Therefore, for any $G \in \mathcal{G}$ we have

$$\hat{\phi} > \lambda(\hat{\phi}) = E[\phi \mid \phi < \hat{\phi}] = \int_{0}^{\hat{\phi}} \phi \ dG(\phi)/G(\hat{\phi}),$$

$$\hat{\phi} < \gamma(\hat{\phi}) = E[\phi \mid \phi > \hat{\phi}] = \int_{\hat{\phi}}^{1} \phi \ dG(\phi)/(1 - G(\hat{\phi})).$$

It necessarily follows that $\lambda(\hat{\phi}) < \hat{\phi}$ and $\gamma(\hat{\phi}) > \hat{\phi}$ for all $G \in \mathcal{G}$.

Define $\hat{\phi}_s$ as the solution to

$$p^*_S(\hat{\phi}_s)(T - 1)[1 - F(\hat{\theta}(p^*_S(\hat{\phi}_s)))] - c_S = p^*_S(\hat{\phi}_a)(1 + \hat{\phi}_a(T - 1))[1 - F(\hat{\theta}(p^*_NS(\hat{\phi}_s)))].$$

And define, $\hat{\phi}_a$ as the solution to

$$p^*_S(\gamma(\hat{\phi}_a))(T - 1)[1 - F(\hat{\theta}(p^*_S(\gamma(\hat{\phi}_a)))] - c_S = p^*_S(\lambda(\hat{\phi}_a))(1 + \hat{\phi}_a(T - 1))[1 - F(\hat{\theta}(p^*_NS(\lambda(\hat{\phi}_a)))).$$

(A-8)
The LHS of these equations are the firm’s average, expected profit given it dispenses samples and the RHS are the firm’s average, expected profit given it does not dispense samples under symmetric and asymmetric information, respectively.

First, because of the monotonicity of the profit functions with respect to \( \mu \), \( \hat{\phi}_s = \hat{\phi}_a \) if and only if \( \lambda(\hat{\phi}_a) = \gamma(\hat{\phi}_a) = \phi \). Second, by letting \( \lambda_a \phi_a = \lambda(\phi_a) \) where \( \lambda_a \in (0, 1) \) and substituting into (A-8) we can totally differentiate to get \( d\phi_a / d\lambda_a > 0 \) for all \( \lambda_a, \phi_a \in (0, 1) \).

Therefore, because \( \hat{\phi}_a \) is monotonically increasing with \( \lambda \), and \( \hat{\phi}_s = \hat{\phi}_a \) if and only if \( \lambda(\hat{\phi}_a) = \gamma(\hat{\phi}_a) = \phi \), it must be the case that the cut-off efficacy level under asymmetric information is less than the cut-off efficacy level under symmetric information for all distributions \( G \in \mathcal{G} \).

**PROOF OF PROPOSITION 4**

When information is asymmetric the sampling decision of the firm partially signals the efficacy to the consumers. The two cases for the proposition follow from the fact that for some values of the parameters there may be no \( \phi \in (0, 1) \) such that \( \Pi^S(\phi) > \Pi^{NS}(\phi) \). We proceed with the proof by assuming that there is a unique cutoff and show that the conditions of the proposition are an equilibrium. We then show that it is the unique perfect Bayesian equilibrium.

Given a unique cut-off \( \hat{\phi} \) such that for all \( \phi \leq \hat{\phi} \) the firm does not dispense samples and for all \( \phi > \hat{\phi} \) it does, the conditional beliefs are

\[
\mu^*_S = E[\phi \mid s = NS] = E[\phi \mid \phi \leq \hat{\phi}] = \int_0^{\hat{\phi}} \phi \, dG(\phi) / G(\hat{\phi}) < \hat{\phi} \quad \text{and} \quad \mu^*_a = E[\phi \mid s = S] = E[\phi \mid \phi > \hat{\phi}] = \int_{\hat{\phi}}^1 \phi \, dG(\phi) / (1 - G(\hat{\phi})) > \hat{\phi}.
\]

In equilibrium the value of \( \hat{\phi}^* \) must represent the firm’s indifference point and satisfy

\[
E\Pi^S(p^*_S(\mu^*_S); \hat{\phi}^*) = E\Pi^{NS}(p^*_S(\mu^*_S); \hat{\phi}^*) = E\Pi^{NS}(p^*_S(\mu^*_S); \hat{\phi}^*).
\]

To show that \( \hat{\phi} < \phi \) where \( \phi \) is the lower bound of the symmetric information sampling set \( \Phi \) we start by assuming \( \hat{\phi} = \phi \). Because \( \mu^*_S > \hat{\phi} = \phi \) we have

\[
E\Pi^S(p^*_S(\phi), \phi) < E\Pi^S(p^*_S(\mu^*_S), \phi).
\]

Likewise, because \( \mu^*_S < \hat{\phi} = \phi \) we have \( E\Pi^{NS}(p^*_S(\phi), \phi) > E\Pi^{NS}(p^*_S(\mu^*_S), \phi) \). Recall that at \( \phi \) the firm is indifferent between dispensing and not dispensing samples under symmetric information; i.e., \( E\Pi^{NS}(p^*_S(\phi), \phi) = E\Pi^S(p^*_S(\phi), \phi) \). Thus it must be the case that \( E\Pi^{NS}(p^*_S(\mu^*_S), \phi) < E\Pi^S(p^*_S(\mu^*_S), \phi) \). \( \hat{\phi} \) is in the asymmetric information sampling set, and \( \hat{\phi}^* < \phi \).

We must now show that when \( \hat{\phi} \) exists, the firm will provide samples for all \( \phi > \hat{\phi} \). To see that it will, consider how the firm’s profit changes with the efficacy of its drug. By the definition of \( \hat{\phi} \), \( E\gamma(\hat{\phi}) = E\Pi^{NS}(\hat{\phi}) \). There are two cases to consider. First when
consumers are *unconstrained* $p_S^*$ is independent of $\mu_S^*$ and it follows from Proposition 1 that for all $\mu_{NS} < 1$, $p_{NS}^*(\mu_{NS}) < p_S^*$. Thus, it is clear that

$$ \frac{dE \Pi^S(\phi)}{d\phi} \geq \frac{dE \Pi^{NS}(\phi)}{d\phi} \quad \forall \phi \geq \hat{\phi}, $$

and it is optimal for the firm to provide samples for all $\phi > \hat{\phi}$.

When consumers are *constrained* the analysis is slightly more involved because $p_S^*$ is belief dependent. Under symmetric information it was shown that the firm will not provide samples for high efficacies. Taking into account when $\hat{\phi} < 1$, the belief regarding the drug’s efficacy when the firm provides samples may alternatively be defined as

$$ \mu_S = \frac{\int_{\hat{\phi}}^{\phi} \phi \ dG(\phi)}{G(\phi) - G(\hat{\phi})}. $$

In this way, the belief regarding the drug’s efficacy when the firm does not provide samples is

$$ \mu_{NS} = \int_{0}^{\hat{\phi}} \phi \ dG(\phi) + \int_{\hat{\phi}}^{1} \phi \ dG(\phi) / 1 - G(\hat{\phi}). $$

If $\mu_{NS} > \mu_S^*$ then the firm will not provide samples for any efficacy because not providing samples induces a higher belief. In fact, in such a case all consumers can infer from a lack of sample dispensation is the unconditional prior. Suppose that $\mu_{NS}^* > \mu_S^*$. Let

$$ \gamma_{NS} = \frac{1 + \mu_{NS}^* (\Sigma - 1)}{\mu_{NS}^* \Sigma}, $$

$$ \gamma_S = \frac{1 + \mu_S^* (\Sigma - 1)}{\mu_S^* \Sigma}. $$

Note that $\mu_{NS}^* > \mu_S^* \Rightarrow \gamma_{NS} < \gamma_S$. From the proof of Proposition 1, since consumers are *constrained* we have established that the firm will cover the same proportion of the market whether it dispenses samples or not. Therefore we have

$$ \theta^* - \gamma_{NS} (p_{NS}^* + \bar{u}(\theta^*)) = \theta^* - \left( [(\Sigma - 1)/\Sigma] p_S^* + \gamma^S \bar{u}(\theta^*) \right) \quad (A-9) $$

Rearranging (A-9) yields

$$ p_{NS}^* = \frac{\mu_{NS}^* (\Sigma - 1)}{1 + \mu_{NS}^* (\Sigma - 1)} p_S^* + \bar{u}(\theta^*) \left( \frac{\gamma_S}{\gamma_{NS}} - 1 \right). $$

Therefore $p_{NS}^* < p_S^*$ and the firm will prefer to dispense samples at $\mu_{NS}^*$ and beliefs are inconsistent.

Thus the sampling decision is monotonic in $\phi$ and $\hat{\phi}$ is the unique cut-off efficacy level. Finally, when $\Pi_S < \Pi^{NS}$ for all $\theta \in [0, 1]$ the firm will not find it optimal to dispense samples for any efficacy. Consequently, consumers will hold the unconditional prior $\mu = \mu_0$.
\[ E\phi = \int_0^1 \phi \, dG(\phi). \] The price the firm sets when it does not provide samples was derived in the text in Section 1.3.

**PROOF OF PROPOSITION 5**

There are two parts to the proposition. Let \( \phi \) be the efficacy of the firm’s drug and let \( \Phi \) be the sampling set of efficacies.

1.) When \( \phi \in \Phi \), then condition (1.21) of the proposition is a result of rearranging (1.20) to find when the resulting expression is greater than 0.

2.) When \( \phi \not\in \Phi \), consumer welfare is still increased because the conditional equilibrium belief given the firm does not dispense samples is less than the unconditional equilibrium belief when the firm is not permitted to provide samples, \( \mu_{NS} < \mu_0 \), thus the optimal no sampling price is less when sampling is permitted.

**PROOF OF PROPOSITION 6**

Let \( \phi \) be the efficacy of the firm’s drug and let \( \Phi \) be the sampling set of efficacies. If \( \phi \in \Phi \), then the result is found by determining when the summation of Eqs. (1.20) and (1.22) are greater than 0.
Appendix B: Chapter 2 Proofs

The following 4 Lemmas characterize further the regulator’s problem for a pure output maximizing firm. First, recall that the firm’s optimization problem is defined as

\[
\max_x \beta \Pi(x; p, T, \theta) + (1 - \beta) \phi(x) \quad \text{subject to} \quad \Pi \geq 0.
\]  

(B-1)

Let \( l(\theta) \) denote the Lagrange multiplier then the firm’s FOCs yield:

\[
\begin{align*}
l &= \frac{(\beta(p - gx) + (1 - \beta)\phi')}{(gx - p)}, \\
px - g + T &= 0, \\
0 \leq l &\leq \infty.
\end{align*}
\]

(B-2)  

(B-3)  

(B-4)

The multiplier, \( l \), identifies the shadow price of increasing firm profit in terms of lost community benefit. It is straightforward to show that the shadow price is decreasing in the cost-state \( l_{\theta} < 0 \), which follows because the firm’s marginal cost of production is increasing with the cost state, making it more expensive to produce the same quantity. Lastly, when, at the limit, the unit price equals the quality-adjusted marginal cost \( p = gx \), the Lagrange multiplier will assume the value \( \infty \). It is clear that when the firm’s optimal choice of output leaves it with some positive profit, \( \beta(p - gx) + (1 - \beta)\phi' = 0 \) and \( l(\theta) = 0 \).

Throughout the analysis we have used the quality-adjusted cost function \( g(x; p, \theta) = c(x, q) \). Continuing to use \( g(\cdot) \), it is useful to denote \( AC_{qa} \) as the quality-adjusted average cost and \( MC_{qa} \) as the quality-adjusted marginal cost, formally

\[
\begin{align*}
AC_{qa} &= g(x, \theta)/x, \\
MC_{qa} &= g_x(x, \theta).
\end{align*}
\]

Because the cost of production is monotonically increasing and convex in \( x \) it is easy to show the following lemma.

**Lemma 7.** There exists a unique \( x \geq 0 \) such that \( AC_{qa}(x) = MC_{qa}(x) \).

**Proof.** This follows from the fact that there exists an \( x \) such that \( AC(x) = MC(x) \) for all \( q \geq 0 \) and the \( MC_{qa} (g_x) \) is increasing convex in \( x \).
Let $x^E$ denote the unique $x$ satisfying $AC_{qa}(x) = MC_{qa}(x)$ and let $x^{fb}$ be the first-best quantity, then inducing the first-best outcome can be characterized by the relative value of $x^{fb}$ to $x^E$.

The following lemmas characterize the first-best policy.

**Lemma 8.** When the firm is a pure output-maximizer, a policy inducing the first-best must include a positive lump-sum transfer for all quantities $0 < x^{fb} < x^E$ and $p(\theta) \leq MC_{qa}(x^{fb}(\theta); p(\theta), \theta)$.

**Proof.** When $x^{fb} < x^E$ then $MC_{qa}(x^{fb}) > AC_{qa}(x^{fb})$. If $T(\theta) = 0$ for all $\theta$ then $p(\theta) \leq AC_{qa}(x^{fb}; p(\theta), \theta)$ if the firm is to produce $x^{fb}$. However, $p(\theta) \geq AC_{qa}(x^{fb}; p(\theta), \theta) \Rightarrow p(\theta) \geq MC_{qa}(x^{fb}; p(\theta), \theta)$ and the firm can continue to produce beyond $x^{fb}$. A positive transfer ($T(\theta) = 0$) functions as a subsidy, lowering the firm’s average cost curve and sliding the efficient scale down the marginal cost curve. The transfer must be sufficiently large to insure $p(\theta) \leq MC_{qa}(x^{fb}(\theta); p(\theta), \theta)$.

**Lemma 9.** When the firm is a pure output-maximizer, the first-best can be induced with only a unit payment, $p$, when $x^E \leq x^{fb}$.

**Proof.** When $x^{fb} > x^E$, it must be the case that $MC_{qa}(x^{fb}) < AC_{qa}(x^{fb})$, by definition of $x^E$. Therefore at $p(\theta) = AC_{qa}(x^{fb}; p(\theta), \theta)$ the firm will increase quality until output equals $x^*$.

Lemma 9 corresponds with lemma 3.6 in Rogerson (1994).

**Lemma 10.** First-best can be induced with only a lump-sum transfer, $T$, for any $x^{fb} > 0$.

**Proof.** This follows immediately from (B-3). Because the firm’s cost is increasing in output, the regulator can induce the firm to output the first-best quantity simply by giving it a lump-sum payment equivalent to the unique cost of producing that output.

**PROOF OF PROPOSITION 8**

For any given $p$, the firm chooses the $x^*$ that sets $p = g_x - ((1 - \beta)/\beta)\phi'$. Thus, the socially preferred output is the output setting $p = V_x - ((1 - \beta)/\beta)\phi'$. Because of the concavity of the regulator’s problem, if $p < V_x - ((1 - \beta)/\beta)\phi'$ for any $p$, then $x^* < \arg \max_x (V - g)$, and if $p > V_x - ((1 - \beta)/\beta)\phi'$ for any $p$, then $x^* > \arg \max_x (V - g)$.

**PROOF OF LEMMA 4**

Condition (i) is a necessary condition for an optimum and is derived as follows. First we derive the condition when $\Pi(x^*) > 0$ and the firm has a mixed objective.

Recall the firm’s value function is defined as

$$U(\hat{\theta}, \theta) = \beta [p(\hat{\theta})x^*(p(\hat{\theta}), \theta) - g(x^*(p(\hat{\theta}), \theta), \theta) + T(\theta)] + (1 - \beta)\phi(x^*(p(\hat{\theta}), \theta)). \quad (B-5)$$

174
A necessary condition for truth-telling is that the announcement of $\theta$ results in maximal profit. The first-order condition for truth-telling is thus

$$
\frac{\partial U}{\partial \theta}(\hat{\theta}, \theta) = \beta \left[ \frac{dp}{d\theta} x^*(p(\hat{\theta}), \theta) + p(\hat{\theta}) \frac{\partial x^*}{\partial p} \frac{dp}{d\theta} - \frac{\partial g}{\partial x^*} \frac{dp}{d\theta} - \frac{\partial g}{\partial p} \frac{dp}{d\theta} + \frac{dT}{d\theta} \right] + (1 - \beta) \frac{d\phi}{dx} \frac{dx^*}{d\theta} \frac{dp}{d\theta} = 0.
$$

Applying the envelope theorem to the first-order condition of $U(\theta) = U(\theta, \theta)$ implies

$$
\frac{dU}{d\theta} \bigg|_{\hat{\theta}=\theta} = \partial U \partial \theta = \beta \left[ \frac{dp}{d\theta} x^*(p(\hat{\theta}), \theta) + p(\hat{\theta}) \frac{\partial x^*}{\partial p} \frac{dp}{d\theta} - \frac{\partial g}{\partial x^*} \frac{dp}{d\theta} + \frac{dT}{d\theta} \right] + (1 - \beta) \frac{d\phi}{dx} \frac{dx^*}{d\theta} \frac{dp}{d\theta}.
$$

where $\hat{\theta}(\theta)$ is the firm's announcement strategy given the true demand state is $\theta$; i.e. $\hat{\theta} : \Theta \rightarrow \Theta$. Thus, by applying the envelope theorem, a necessary condition for the optimal payment policy is that

$$
\frac{dU}{d\theta} = -\beta \frac{\partial g}{\partial \theta}.
$$

When $\Pi(x^*) = 0$ and the firm is a pure output-maximizer then $x^*$ solves $px^* - g(x^*; p, \theta) + T = 0$. Because the firm leaves itself with 0 profits for any state $\theta$, we have

$$
\frac{dU}{d\theta} = \beta \frac{d\Pi}{d\theta} + (1 - \beta) \frac{d\phi}{dx} \frac{dx^*}{d\theta} = (1 - \beta) \frac{d\phi}{dx} \frac{dx^*}{d\theta}.
$$

Using the implicit function theorem, $\frac{dx^*}{d\theta}$ can be expressed as

$$
\frac{dx^*}{d\theta} = -\frac{p_\theta \frac{df}{d\theta} x - g_\theta p_\theta \frac{df}{d\theta} - g_\theta + T_\theta \frac{df}{d\theta}}{p - g_x} = \frac{dx^*}{d\theta} \frac{d\hat{\theta}}{d\theta} - \frac{g_\theta}{p - g_x}.
$$

The firm announces the $\hat{\theta}$ which maximizes output, therefore by the envelope theorem the first term on the RHS of (B-7) is zero, yielding:

$$
\frac{dU}{d\theta} = \partial U \partial \theta = -(1 - \beta) \phi'(x^*(p(\theta), \theta)) \left[ \frac{g_\theta(x^*(p(\theta), \theta); p(\theta), \theta)}{g_x(x^*(p(\theta), \theta); p(\theta), \theta) - p(\theta)} \right].
$$

Next, condition (ii) of the lemma represents a sufficient condition. To show sufficiency when the firm's choice of output leaves it with positive profit, we apply the envelope theorem to (B-5) yielding:

$$
\frac{\partial U(\hat{\theta}, \theta)}{\partial \theta} = \beta \left[ \frac{dp}{d\theta} x - \frac{\partial g}{\partial p} \frac{dp}{d\theta} + \frac{dT}{d\theta} \right].
$$

From the fact that $\frac{\partial U(\hat{\theta}, \theta)}{\partial \theta} \bigg|_{\hat{\theta}=\theta} = 0$ we have

$$
\frac{dT}{d\theta} = \frac{\partial g}{\partial p} \frac{dp}{d\theta} - \frac{dp}{d\theta} x(p(\hat{\theta}), \hat{\theta}).
$$
Plugging (B-9) into (B-8) yields
\[
\frac{\partial U(\hat{\theta}, \theta)}{\partial \theta} = \beta \frac{dp}{d\theta} \left[ (x(p(\hat{\theta}), \theta) - g_p(x(p(\hat{\theta}), \theta); p(\hat{\theta}, \theta)) - (x(p(\hat{\theta}), \theta) - g_p(x(p(\hat{\theta}), \hat{\theta}; p(\hat{\theta}, \hat{\theta})) \right]
= \beta \frac{dp}{d\theta} [U_p(\hat{\theta}, \theta) - U_p(\hat{\theta}, \hat{\theta})].
\]

By the intermediate value theorem there exists a \( \hat{\theta} \in [\theta, \hat{\theta}] \) if \( \theta < \hat{\theta} \) or \( \hat{\theta} \in [\hat{\theta}, \theta] \) if \( \theta > \hat{\theta} \) such that
\[
\frac{\partial U(\hat{\theta}, \theta)}{\partial \theta} = \beta \left[ \frac{dp}{d\theta} \frac{\partial^2 U(\hat{\theta}, \hat{\theta})}{\partial \theta \partial \theta} (\theta - \hat{\theta}) \right].
\]
(B-10)

Because \( U_T = \beta \), the second-order cross partial derivative of Eq. (B-10), \( U_{\theta p} \), is equal to \( \frac{\partial}{\partial \theta} (U_p/U_T) \). The condition \( \text{sign} \{ dp/d\theta \} = \text{sign} \left[ \frac{\partial}{\partial \theta} (U_p/U_T) \right] \) implies
\[
\frac{\partial U(\hat{\theta}, \theta)}{\partial \theta} \geq 0 \text{ when } \hat{\theta} < \theta \\
\frac{\partial U(\hat{\theta}, \theta)}{\partial \theta} \leq 0 \text{ when } \hat{\theta} > \theta
\]

Thus, \( \hat{\theta} = \theta \) is a global maximizer and the payment policy induces truthful revelation if \( \text{sign} \{ dp/d\theta \} = \text{sign} \left[ \frac{\partial}{\partial \theta} (U_p/U_T) \right] \).

To prove condition (ii) for a pure output-maximizing firm we employ a slightly different approach. Incentive compatibility is satisfied if and only if the firm maximizes its output with a truthful announcement of the state. Therefore incentive compatibility is satisfied for an output-maximizing firm if and only if for any \( \theta_1 \) and \( \theta_2 \) in \( \Theta \) where \( \theta_1 < \theta_2 \), the following hold
\[
x^*(p(\theta_2), T(\theta_2), \theta_1) \leq x^*(p(\theta_1), T(\theta_1), \theta_1), \quad (B-11)
\]
\[
x^*(p(\theta_1), T(\theta_1), \theta_2) \leq x^*(p(\theta_2), T(\theta_2), \theta_2). \quad (B-12)
\]

Adding (B-11) and (B-12) gives
\[
x^*(p(\theta_2), T(\theta_2), \theta_2) - x^*(p(\theta_1), T(\theta_1), \theta_2) \geq x^*(p(\theta_2), T(\theta_2), \theta_1) - x^*(p(\theta_1), T(\theta_1), \theta_1),
\]
implying
\[
\int_{\theta_1}^{\theta_2} \int_{\theta_1}^{\theta_2} \frac{\partial^2 x^*}{\partial \theta \partial \theta} d\theta d\theta \geq 0. \quad (B-13)
\]

Because (B-13) is true for all \( \theta_1, \theta_2 \in \Theta \) it implies \( \frac{\partial^2 x^*}{\partial \theta \partial \theta} \geq 0 \), which is equivalent to
\[
\frac{\partial^2 x^*}{\partial T \partial \theta} \frac{dT}{d\theta} + \frac{\partial^2 x^*}{\partial p \partial \theta} \frac{dp}{d\theta} \geq 0. \quad (B-14)
\]

We can simplify (B-14) by observing that a truthful announcement of the state parameter is optimal if
\[
\frac{dx^*}{d\theta} \bigg|_{\theta=\theta} = \frac{dx^*}{dp} \bigg|_{\theta=\theta} \frac{dp}{d\theta} \bigg|_{\theta=\theta} + \frac{dx^*}{dT} \bigg|_{\theta=\theta} \frac{dT}{d\theta} \bigg|_{\theta=\theta} = 0. \quad (B-15)
\]
Using (B-15) we can rewrite (B-14) as

\[
\frac{\partial}{\partial \theta} \left( \frac{dx^*/dp}{dx^*/dT} \right) \left. \frac{dp}{dT} \right|_{\hat{\theta} = \theta} \geq 0. \tag{B-16}
\]

Eq. (B-16) is a special case of the condition derived in Theorem 1 of Guesnerie and Laffont (1984). The term \((dx^*/dp)/(dx^*/dT) = (dU/dp)/(dU/dT)\) is the output-maximizing firm’s MRS of unit payment for fixed transfer so the firm’s objective function satisfies the SCP when \(d{(dU/dp)/(dU/dT)} / d\theta\) is monotonic for all \(\theta \in \Theta\).

PROOF OF COROLLARY 2

Starting with the definition of \(U(\theta)\) and property \((i)\) of Lemma 4 we have:

\[
U(\theta) = U(\bar{\theta}) + \beta \int_{\theta}^{\bar{\theta}} g_\theta(x^*; p, \theta)d\theta = (1 - \beta)\varphi(\bar{\theta}) + \beta \int_{\theta}^{\bar{\theta}} g_\theta(x^*; p, \theta)d\theta. \tag{B-17}
\]

Substituting \(U(\theta) = \beta \Pi(\theta) + (1 - \beta)\varphi(\theta)\) in (B-17), rearranging, and taking expectations gives:

\[
\mathbb{E}\Pi = \int_{\theta}^{\bar{\theta}} \left\{ \frac{1-\beta}{\beta} (\varphi(\bar{\theta}) - \varphi(\theta)) + \int_{\theta}^{\bar{\theta}} g_\theta(x^*; p, \theta)d\theta \right\} dF(\theta).
\]

Integrating by parts yields the expression:

\[
\mathbb{E}\Pi = \int_{\theta}^{\bar{\theta}} \left\{ \frac{1-\beta}{\beta} (\varphi(\bar{\theta}) - \varphi(\theta)) - \frac{F(\theta)}{f(\bar{\theta})} g_\theta(x^*; p, \theta) \right\} dF(\theta).
\]

PROOF OF PROPOSITION 9

Because the regulator’s problem is quasiconcave in \(x\) and \(p\) (see footnote 64) we have \(d^2\{V - g\}(x; p) = d^2\{V - g\}(p)/dp^2 < 0\). The first-best price, \(p_{fb}^{fb}\) is the price solving \(d\{g_\theta(x^*; p, \theta)\} / dp = 0\). When \(d\{g_\theta(x^*; p, \theta)\} / dp = 0\) for all \(\theta \in \Theta\) the firm extracts no rents and from Eq. (2.12) it is clear that \(p_{sb}^{sb}\) solves \(d\{V - g\} / dp = 0\). Therefore \(p_{sb}^{sb} = p_{fb}^{fb}\). When \(d\{V - g\}(p_{sb}^{sb}) / dp > 0\) concavity in the regulator’s problem implies \(p_{sb}^{sb} < p_{fb}^{fb}\) and when \(d\{V - g\}(p_{sb}^{sb}) / dp < 0\) concavity implies \(p_{sb}^{sb} > p_{fb}^{fb}\). From Eq. (2.12), \(\text{sign}\{d\{V - g\}(p_{sb}^{sb}) / dp\}\) depends on whether the firm’s information rents are increasing or decreasing with the unit payment. Therefore, when the firm’s information rents are decreasing with the unit payment then \(d\{g_\theta(x^*; p, \theta)\} / dp < 0\) and we have \(p_{sb}^{sb} > p_{fb}^{fb}\) and

\[123\text{This condition can equivalently be written as}\]

\[
\frac{\partial}{\partial \theta} \left( \frac{\partial x^*/\partial T}{\partial x^*/\partial p} \right) \left. \frac{dT}{d\theta} \right|_{\hat{\theta} = \theta} \geq 0.
\]

177
when the firm’s information rents are increasing then \( d\{g_\theta(x^*:p,\theta)\} / dp > 0 \) and we have \( p^{sh} < p^{fb} \).

PROOF OF PROPOSITION 11

The conditions of Lemma 4 identify the necessary and sufficient conditions for an incentive compatible payment policy that must be satisfied by any set of payment rules: \( \{p^{np}(\theta),T^{np}(\theta)\}_\theta \). The unit payment that induces the first-best level of output is derived by setting \( dx^*/d\theta = dx^{fb}/d\theta \), where \( dx^*/d\theta \) is defined in Lemma 4 and solving for \( p \). By the integral form of the envelope theorem, we have \( x^*(p(\theta),\theta) = x^{fb}(\theta) - \int_\theta^\theta (\partial x^{fb}/d\theta) d\theta \), if and only if \( x^* = x^{fb} \) at every \( \theta \in \Theta \), thus a payment policy that induces \( dx^*/d\theta = dx^{fb}/d\theta \) and sets \( x^*(\theta) = x^{fb}(\theta) \) induces the first-best output in every state.

From Lemma 8, the unit payment must always be less than or equal to the marginal cost at the induced quantity. Because \( g_\theta > 0 \), this condition requires that the first-best output be weakly decreasing with the cost-state, \( dx^{fb}/d\theta \leq 0 \), otherwise, \( p^{np}(\theta) > g_x(x^*(\theta);p^{np}(\theta),\theta) \) and the firm is not left with zero profit. Consequently it will choose a higher output and the payment rule does not induce the first-best output. It must also be the case that \( p(\theta) \neq g_x(x^*(\theta);p(\theta),\theta) \) for all \( \theta \in \Theta \), otherwise \( dx^*/d\theta \) is not bounded at some \( \theta \in \Theta \), hence not absolutely continuous, and the envelope theorem cannot apply. The assumptions on the value and cost functions insure this cannot happen. First, the regulator’s problem insures that \( p^{np}(\theta) \) is unique for every state \( \theta \); and second, the firm’s problem insures that \( x^* \) is unique for every \( p \). Therefore, for a given state the first-best quantity is unique and there does not exist any such \( \theta' \in \Theta \) such that \( \lim_{\theta \rightarrow \theta'} dx^{fb}(\theta)/d\theta = \infty \).

Finally, the unit payment must be restricted to a nonnegative value because demand is not defined outside the domain \( p \geq 0 \) and the pricing rule does not guarantee that the price will be nonnegative.

PROOF OF PROPOSITION 12

The regulator’s problem is complicated by the fact that, in addition to the firm’s choice of output, the consumers’ demand is also a function of the unit price, \( p \). Inducing the consumers to demand the first-best quantity requires that \( p^{np} \) solve

\[
\frac{d\{V - p\}}{dp} = 0. \tag{B-18}
\]

Inducing the firm to supply the appropriate level of quality while maintaining incentive compatibility requires that \( p^{np} \) also satisfy

\[
p^{np}(\theta) = g_x(x^{fb};p^{np}(\theta),\theta) + \frac{g_\theta(x^{fb};p^{np}(\theta),\theta)}{dx^{fb}/d\theta} \quad \text{for all } \theta \in \Theta. \tag{B-19}
\]

Eq. (B-18) and (B-19) are independent, thus \( p^{np}(\theta) = p^{fb}(\theta) \) for all \( \theta \in \Theta \) only if \( dp^{np}/d\theta = dp^{fb}/d\theta \) at all \( \theta \in \Theta \). It is clear that, generically,

\[
\frac{dp^{fb}}{d\theta} = -\left( \frac{\partial^2 \{V - g\}}{\partial p \partial \theta} / \frac{d^2 \{V - g\}}{dp^2} \right) \neq \frac{dp^{np}}{d\theta}.
\]

178
Hence, \( p^{np} \neq p^f_b \) and Eq. (B-18) is not equal to zero at all \( \{ x^{fb}(\theta), p^{np}(\theta), T^{np}(\theta) \}_{\theta \in \Theta} \) given the strict concavity of the regulator’s problem. Therefore \( \{ x^{fb}(\theta), p^{np}(\theta), T^{np}(\theta) \}_{\theta \in \Theta} \) is not first-best optimal for a marketed good.

**PROOF OF LEMMA 6**

The proof for Lemma 4 continues to hold if \( g_p = 0 \) so applies to a nonmarketed good as well. For a nonmarketed good the SCP holds and is negative for all class of functions satisfying the model’s properties. Therefore the condition \( \text{sign} \left[ \frac{dp}{d\theta} \right] = \text{sign} \left[ \frac{\partial}{\partial \theta} \left( \frac{U_p}{U_T} \right) \right] \) requires \( \text{sign} \left[ \frac{dp}{d\theta} \right] < 0 \).

**PROOF OF PROPOSITION 13**

Combining Lemmas 8 - 10 yields the pricing rule of the proposition.

**PROOF OF PROPOSITION 15**

The FOC of the Hamiltonian must satisfy:

\[
\frac{d}{dp} \left\{ \frac{d}{d\theta} \left( V - g \right) \right\} = \frac{d}{dp} \left\{ (V_x - g_x) \frac{dx^*}{d\theta} \right\} = 0, \\
= \frac{d}{d\theta} \left\{ (V_x - g_x) \frac{dx^*}{d\theta} \right\} = 0.
\]

Where \( V_p = g_p = 0 \) since there is no demand response to price. The FOC can be manipulated to isolate \( \frac{dx^*}{d\theta} \); that is, rearranging the FOC yields:

\[
\frac{dx^*}{d\theta} = -\frac{d}{d\theta} \left\{ (V_x - g_x) \frac{dx^*}{d\theta} \right\} = -\frac{d}{dx} \left\{ \frac{d}{dp} (V - g) \right\}.
\]

The RHS of (B-20) is the definition of \( \frac{dx^{fb}}{d\theta} \), therefore if the payment policy induces \( \frac{dx^*}{d\theta} = \frac{dx^{fb}}{d\theta} \) then the first-order necessary condition for maximization is satisfied. Because the regulator’s problem is quasi-concave the condition is also sufficient and the regulator achieves the first-best outcome.
Bibliography


