Firm Pricing and Entry

by

Fiona M. Scott Morton

B.A., Economics, Yale University 1989

Submitted to the Department of Economics
in Partial Fulfillment
of the Requirement for the Degree of

DOCTOR OF PHILOSOPHY
in Economics
at the

Massachusetts Institute of Technology

April 1994

Signature of Author . . . . . .
Department of Economics
May 12, 1994

Certified by . . . . . .
Jerry A. Hausman
John and Jennie S. MacDonald Professor of Economics
Thesis Advisor

Certified by . . . . . .
Nancy L. Rose
Professor of Economics, Sloan School of Management
Thesis Advisor

Accepted by . . . . . .
Olivier J. Blanchard
Professor of Economics

Firm Pricing and Entry

by

Fiona M. Scott Morton

Submitted to the Department of Economics on April 8, 1994
in partial fulfillment of the requirements for the
Degree of Doctor of Philosophy in Economics

Thesis Advisors: Professor Jerry A. Hausman, MacDonald Professor of Economics
Professor Nancy L. Rose, Sloan School of Management

Abstract

Chapter 1: Modern examples of situations conducive to predatory pricing are scarce, whereas international shipping at the turn of the century was subject to little, if any, antitrust regulation. Shipping firms often formed cartels on international routes in the late nineteenth century and into the twentieth century. A cartel determined schedules and fixed prices for its member firms. This chapter examines the response of three cartels to entry of independent shipping firms onto the cartelized route. The cartels in my sample sometimes reacted to entry by initiating a price war. Many models of predatory pricing demonstrate that an incumbent has an incentive to prey on a weak entrant rather than a strong one. I find that the cartels tended to start price wars when the entrant was weak in one of several ways. A weak entrant had less shipping experience, less accumulated resources, no contract for cargo on the route, and unfavorable trade conditions. I conclude that the behavior of early shipping firms supports the hypothesis that the benefit to predation varies with the characteristics of the entrant.

Chapter 2: Entry in the pharmaceutical industry can take the form of a branded product or a generic product. A product which loses patent protection may attract zero or many generic entrants. I develop a Poisson count model of entry that analyzes the characteristics of a market which attracts generic entrants. I find that the revenue of the patented drug before it loses patent protection and the FDA's regulatory environment are the most important factors in determining the number of generic entrants. I then discuss the role of advertising by the brand firm to physicians. If advertising enlarges the perceived quality difference between the brand and generic drug, it could be a barrier to entry on the part of generic firms. Following the previous literature, I include advertising as an explanatory variable in the entry equation and find it has a very small, negative coefficient. However, advertising is not exogenous to entry. I therefore instrument for advertising and find that its coefficient doubles and remains negative and significant (the expected result if the endogenous aspect of the correlation is removed). However, the coefficient is not economically significant; its low level indicates that deterring generic entry through brand advertising is very expensive.

Chapter 3: I examine the response of pharmaceutical firms to the Most Favored Nation Clause included in the Medicaid Rebate Law that took effect in January 1991. One of the parts of the law required that the price Medicaid paid for a *brand name* prescription drug would be the lowest price offered on that drug to any other drugstore. I compare the response of mean prices and quantities in a sample of cardiovascular drugs. Because the MFN clause did not apply to drugs sold to hospitals or to generic drugs, I can use them as control groups. I find mean prices rose as a result of the law. I also look for a strategic response on the part of generic drugs in a market with a brand. In markets with only one or two generic manufacturers, the generic oligopolist should respond to the brand's constraint by raising prices; in markets with many generic manufacturers, the brand has negligible impact compared to all the other generic firms. I do not find evidence supporting the existence of the strategic effect.

Acknowledgements

I am grateful for financial support from the MIT Economics Department, the Sloan Foundation, by way of the Program on the Pharmaceutical Industry at the MIT Sloan School, and the George and O'Bie Shultz Fund.

This thesis is not a product of one person, but of many. First and principally, many classmates and officemates provided help and camaraderie over the years: Yacine Ait-Sahalia, Maura Doyle, Glenn Ellison, Sara Fisher Ellison, Michael Gordy, Austan Goolsbee, Ruth Judson, Peter Klibanoff, David Laibson, Owen Lamont, Alex Pfaff, Jack Porter, Tom Skinner, Matt Slaughter, Chris Snyder, Winnie Yip, and Jeromin Zettlemeyer, to name just a few.

Without my housemates, Margaret Chen, Mary Hallward, and Amy Hiestand, life outside of MIT would not have been as full of fun and humor.

I would especially like to thank my two friends and classmates, Judy Chevalier and Brigitte Madrian, without whom I undoubtedly would not have finished all my problem sets nor enjoyed my years at MIT so much.

Many faculty at MIT, Frank Fisher, David Genesove, Whitney Newey, Danny Quah, and Jean Tirole among them, gave time and advice which I appreciate.

I am especially grateful to my advisors, Jerry Hausman and Nancy Rose, for helping with endless empirical problems and generously giving me their time and the benefits of their experience. I could not have asked for better advisors.

Thanks also to Stephen Latham for his invaluable emotional support and editing talents.

Finally, I would like to thank the members of my family for their continuing understanding, enthusiasm, and assistance while I was working on this degree.

Table of Contents

Introduction

Each of the three chapters below provides empirical evidence relating to a particular theoretical controversy in the field of Industrial Organization. The theoretical side of I.O. has advanced enormously in the last several decades and empirical work has been struggling to keep up. The goal of these three studies is to improve our understanding of firm behavior so that antitrust laws, regulations, and other policy decisions that have an impact on firms can be designed well and improve social welfare. Datasets which can be used to study firm behavior are much more scarce and less comprehensive than the datasets that have long been available in the Public Finance and Labor fields. Often, data such as investment spending or costs are badly measured or not reported at all. As a result, a researcher has to turn to measures which are easier to detect or regulate. Entry is one such measure. All three chapters take advantage of the information contained in a firm's decision to enter a particular market.

The first chapter examines the much-studied area of predatory pricing. The theoretical literature includes both work that claims it can never be logical for a firm to undertake predatory pricing and work that concludes that under some circumstances, a firm should always prey. The shipping industry at the turn of the century provides an excellent place to look for evidence because there was essentially no antitrust law at the time. I find that shipping cartels did prey occasionally; particularly if the entrant was weak in one of several dimensions such as resources, cargo contracts, or cyclical trade movements.

The second chapter switches both industry and time period, examining the U.S.

pharmaceutical industry of the late 1980s. Again, the issue of whether a barrier to entry can exist is hotly debated in the theoretical literature. Advertising is often postulated to be such a barrier. I examine recently expired pharmaceutical patents and look at whether advertising to doctors discourages generic entry. Regulation by the FDA means that pharmaceutical markets are very clearly defined, as are entry dates and the identity of the entrant. I discover that advertising to doctors seems to discourage entry, but that the effect is very small. A firm would have to spend a prohibitive amount of money to deter a generic entrant, so the advertising should not be a concern to policy makers on that account.

The final chapter looks for an empirical response to the Most Favored Nation clause applied to pharmaceutical prices in 1991. Theory tells us that a MFN clause promotes softer price competition in an oligopoly, but the theory has not been tested against any data. In the case I examine, Medicaid announced a complex MFN rule for its pharmaceutical purchases. The rule varied across firms and markets; I exploit that variation to see if the MFN clause caused price and quantity to shift. I find the average price of a product covered by the MFN clause rises, and those products with generic substitutes lose quantity. The government policy reduced competition in the pharmaceutical market and essentially created an indirect tax on American consumers.

Chapter I

British Shipping Cartels 1879-1929:

Entry and Price Wars

I. Introduction

This paper examines the motivations for price wars which occurred within British merchant shipping cartels in the late 1800's. British shipping cartels operating around the turn of the century provide an excellent example of cartel behavior. The cartels, or "conferences," were associations of international shipping lines which fixed prices and quantities on their routes during the period 1880 to 1929. I consider why we might observe shipping conferences of the period initiating price wars against some lines that tried to enter their routes. One explanation is that the cartels engaged in predatory pricing against these entrant lines. However, other entrants into the same routes in the same time period were not preyed upon. I consider what motivated the cartel's choice of whether or not to begin a price war.

Well-known scholars such as Bork (1978) have argued that predatory pricing is unlikely to be profitable for a firm, much less a group of firms, and will therefore not occur. A firm expecting to gain from predatory pricing must be able to recoup its short run losses by earning excess profits in the long run. Those future profits depend on its successfully driving out the entrant or rival and then maintaining monopoly power long enough to earn back its losses plus some profit. The problem becomes even more difficult should a cartel, rather than a single firm, consider predation. The distribution of losses and gains must be arranged, and free-rider cheating must be controlled.[1] The

[1] Pareto improving options, such as merger with the entrant, would be chosen over such a price war, according to McGee (1983). However, merging with every entrant will *attract* entry and will not be a good long run strategy.

conservative camp concludes that price wars are not evidence of predation, but merely evidence of competition; perhaps cost changes or demand shocks trigger the periods of low prices, but the intent is not predatory. The discussion and estimation in the paper seeks to show that in the shipping industry at the turn of the century, price wars were begun as a premeditated strategy to drive out entrants, and suggests what some of those motivations might have been.

Characteristics of the shipping industry mitigate some of the problems with predatory pricing mentioned above. For example, the incumbent would normally have to buy the capacity of a vanquished rival, or another rival would simply enter. Because ships can ply a wide variety of routes, sunk *capacity* is not an issue. However, a certain amount of infrastructure was necessary to operate a merchant shipping route: brokers, docking rights, and coaling stations were the most important components.[2] Those fixed costs meant that the industry was not contestable. Merging has been advanced as a cheaper alternative than predation[3]; in fact that is what these cartels did in over half the cases of entry I study. In these cases the entering firm was formally admitted to the cartel after negotiations over market share. The question might therefore be restated, why did the cartel undertake price wars in the remaining cases?

The organization and discipline of the shipping cartels aided successful predation. Only a handful of firms participated in a given cartel. The heads of the firms could easily meet and decide prices and quantities. The cartel set not only prices, but also

---

[2] Brokers would organize freight, collect fees and build up a regular customer base.

[3] See footnote 1.

allocated market shares (number of sailings), decided the exact ports to be served, and often enforced some revenue sharing scheme also. Cheating, by sending more than the allowed number of ships on a route, was completely visible to all participants in the market. Deviating by lowering price, when output could not be substantially expanded, would not increase profits. Advertising low rates to fill an allotted ship would likewise be discovered; the shipping world was small and talkative. Under the controls these shipping cartels established, agreed upon prices could often be sustained.

Identifying predatory pricing accurately requires knowledge of variable costs at the very least, and ideally, knowledge of marginal cost. Since I do not have either, it will be impossible for me to prove predatory pricing existed. Future work will focus on this problem as I expect to have access to cost data of several firms. However, to determine sound cases of predatory pricing, Easterbrook (1981) suggests finding evidence of firms being driven out of the market and *then* of prices rising. The data in Table I confirm that prices fell and then rose in situations where the entrant was driven out as well as when it stayed.

However, the behavior and data we are able to observe are consistent with the predatory pricing motive. The signs of the variables which successfully predict price wars support several strategic theories of predatory pricing. I use firm histories to identify entries, cases of price wars, and characteristics of the firms involved. Then I test the entries to determine what causes only some entrants to be fought by the cartel

while others enter without contest.[4] The three cartels examined below were chosen for their cohesiveness and stability; at no time in the period did their collusion mechanism break down or prices float. I find that theoretical motives such as long purse and renegotiation are supported by the price war choices of these shipping cartels. The cost of a price war to the cartel, measured by trade variables, also affects the likelihood that the cartel will initiate one. Entrant-specific characteristics which are unrelated to the state of competition, demand, or supply shocks affect the likelihood of a price war. This result supports the hypothesis that the price wars were motivated by predatory intent and were not simply the outcome of vigorous competition.

The organization of the paper is as follows. In section II I discuss entry theory and a possible model of shipping cartels. Section III describes common practices of the cartels, analyzes the extent of their market power, and gives a rough picture of shipping firm finances.[5] The historical detail of the three cartels I use in my dataset is covered in Section IV. Section V explains the variables used in the estimation and discusses the results. The problem of misclassification of the dependent variable and its solution is

---

[4] The focus here is thus different from Porter (1983). I look at the cartel's response to entry; Porter mainly addresses the possibility of secret price cuts within the cartel. Green and Porter (1983) and Ellison (1991) both use the JEC price data to solve for a switching regime of high prices and punishment phases.

[5] Considerable scholarship has applied the theory of the core to the problem of ocean shipping e.g. W. Sjostrom (1992) and S.J. Pirrong (1992). The basic idea is if units of supply, ships in this case, are big enough relative to the market, there might not be an equilibrium because of integer constraints. In such a case, a cartel may be the most efficient solution for bringing stability to the market. This explanation is offered to account for the persistence and success of the shipping conferences of the period. Although the empty core may well have contributed to cartel formation, persistent quantity limits and other restrictions on competition suggest that the cartel purposefully applied its power to raise prices above average cost.

discussed in Section VI and Section VII concludes.

II. Entry Theory

The stylized story of a cartel threatened by entry and responding with predatory prices is well known. The monopolized market is entered by a new firm, the incumbent drops its price[6], a price war ensues, the entrant is forced into bankruptcy and exits, and finally the market returns to monopoly price and quantity. Alternatively, the cartel decides to allow the entrant to share the market either before or after the price war has begun.

Ordover and Saloner (1987) list three reasons a firm might undertake predatory pricing. The first is asymmetrical finances, or the long purse story, where war begins because each side believes it has the greater financial resources. If the incumbent's superiority were common knowledge, entry would not be an equilibrium strategy for the entrant. Some versions of the theory depend on imperfect capital markets. The well developed capital markets of the nineteenth century shipping industry suggests that imperfections were not be as prevalent as one might assume given the historical period.

> "The boom in the market for stocks and shares of all kinds after the repeal of the Bubble Act in 1825 included massive joint-stock promotions in the shipping industry...The new capital was sorely needed...to cover the much higher costs of iron-hulled steam-powered fleets...and no less than 413 shipping companies were registered under the new limited-liability legislation between 1856 and 1881. The more extensive use of limited liability in the 1890s and early 1900s improved

[6] To exactly what level is a subject of disagreement.

owners' opportunities for introducing external finance."[7]

Limited liability allowed the owners to use external financing more easily. However, Green finds that ship owners had not been eager to borrow from banks in the early part of the period.

> "For most of the nineteenth century the industry had been reluctant to ask banks and other institutions for support...'it was never company policy to attempt to secure outside finance.' For their part banks had been unenthusiastic about closer involvement."[8]

Capital markets were certainly less developed during this period than during the modern age and perhaps more imperfect also; but it remains to be shown that the long purse theory is an important explanation of price wars in the case of shipping conferences.

Signalling (asymmetric information) is a second explanation of predatory price wars. Ordover and Saloner compare signalling to standard limit pricing, except that the entrant has already entered. Saloner (1987) notes that the incumbent is trying both to induce exit and to improve its own position in case the entrant actually stays. For example, a price war might lower acquisition price (McGee 1980) or, by analogy, convince the entrant to accept a lower number of sailings or less desirable ports of call in the final conference agreement.

Creation and defense of a reputation is the third story behind predatory pricing. Selten's chain store paradox (1978) shows that predatory pricing in the face of entry is irrational in a finite game. Milgrom and Roberts (1982) look at reputation and its

---

[7] Green, "Ownership" p229.

[8] Green, "Ownership" p230.

importance in the case of uncertainty. They prove that when potential entrants can enter every period, the incumbent's reputation for preying will have a deterrent effect on entry and therefore some initial predation may be rational. Their result depends on a finite horizon for the game, which is not the case (ex ante) in most industries, including shipping. However, Fudenberg and Levine (*Econometrica* 1989) show that in the case of an infinite horizon, a monopolist committed to a strategy can get at least his Stackelberg equilibrium payoff if he is sufficiently patient. Preying to convince entrants of its strategy can be profitable for a monopolist in a game that does not end with certainty in any period.[9]

Two General Types of Predation Models

The Industrial Organization literature contains many theoretical models in which predatory behavior of some sort is the incumbent firm's optimal strategy. There are two main types of predation models. Those such as Milgrom and Roberts (1982) or McGee (1980) derive very strict results; predation is rational but never necessary, or useless and never used. In constrast, other models predict predatory behavior will be more common or more often optimal when the entrant has better "survivability" characteristics, or is "tougher." These papers allow for an entrant's characteristics to affect the chance of a price war. Whether the entrant's toughness is a long purse or an investment of some kind, the general conclusion we can draw from this strand of the literature is that

[9] Indeed, the incumbent can to do better than Stackelberg, if it is faced with forward-looking entrants, by committing to a strategy over time.

predation is less likely to be undertaken against firms which are stronger. If this sort of model corresponds to the shipping market of the late 1800s, then entrant and market characteristics should help explain the probability of a price war.

For example, Fudenberg and Tirole (1985) propose a theoretical model for the traditional 'long purse' story. In their model, predation causes the entrant to have sufficiently little cash after entry that it cannot stay in the market; no bank will lend to it at a profitable interest rate. An entrant with more cash, or a better relationship with its bank will attract less predation. In an extension of the Fudenberg and Tirole model, Snyder (1993) shows that long term contracts between an entrant and its bank can succeed as a predation defense in some cases. The long term nature of the financing strengthens the entrant. Bolton and Sharfstein (1990) examine yet another 'long purse' type of case. Suppose a firm is financially constrained in order to provide managers with incentives, then its rivals have a motive for attempting to make the firm earn losses in a price war and exit. The price war will only be successful if the entrant is sufficiently weak that the incumbents can force exit.

Benoit (1983) discusses a model where predation is threatened and notes that in a situation of perfect information, no entry ever occurs. However, if incomplete information is introduced (Benoit (1984)), some entry is undertaken, and those entrants which find exit most costly are successful. In my data, both the shipping entrant and the cartel would have had incomplete information concerning each other. Most of the shipping entrants were private firms whose capital was owned by one or two extended families and neighbors. Each would have a different mixture of loans and equity, and

partners with different wealth and liquidities. Thus the cartel could not have known the exact characteristics of its entrants. Also, the entrant would not have known the attributes of the market as well as the cartel did and would thus make an imperfect estimate of its true strength.

Ordover and Saloner (1987) suggest signalling as a motivation for predation. Under this hypothesis, predation will only occur in cases where the entrant lacks information. Similarly, in signalling models like Fudenberg and Tirole (1986 'Signal Jamming'), predation is used when it the entrant is less informed or weaker, and can be convinced to exit. The type of model which allows for different outcomes depending on entrant and market characteristics motivates the empirical work that follows; I look for variables which affect the probability of a price war.

III. Shipping Cartels

Ocean-going freight may be shipped in a liner or in a tramp. The distinction is not a physical one, rather it depends on how the ship is scheduled. Tramps are paid to pick up a full shipload of cargo X at port A and carry it to port B; they operate on a spot market where rates may fluctuate overnight according to supply and demand. Liners, on the other hand, commit to a regular schedule of departures and stops at intermediate ports no matter how much cargo may be there to carry. The cost structure of liner shipping is similar to other scheduled transportation systems: high fixed costs and very low marginal costs (handling of cargo) up to capacity. Therefore freight shippers have a powerful incentive to take on cargo at marginal cost if the ship is not full; obviously,

they will not cover fixed costs if all cargo pays such a rate.

Shipping Conferences are associations of deep sea merchant or passenger shipping lines. Their origins trace to the 1870s when steamships were just technologically advanced enough to travel the world's longest trade routes. In the mid to late nineteenth century, sailing ships were still playing an important role in world trade, but the steamship was overtaking rapidly. At the time, British shipping companies carried over 50% of world trade and owned a similar percentage of world steam tonnage.[10] The British lines were the largest, most technologically advanced, and dominant in most trade routes during the period. Additionally, British lines were privately held and did not receive large government subsidies, providing a better experiment for the economist.[11] (However, substantial UK mail subsidies did exist.)

A Shipping Conference is composed of a number of lines, all travelling the same route. A line may be quite small, three ships for example, or could be large enough to cover many routes all over the world, as many of the well-known shipping names did. The original purpose of the shipping conferences was to set rates and sailing schedules to which each line would adhere. Each conference agreement would apply to only one direction of one route, London to Calcutta for example, because the opposite direction might be served by a different set of firms. The conference agreement would include all lines travelling the route at that time and might have from two to fifteen members. Two

---

[10] D. Aldcroft (1968)

[11] Later in the century (after WWII primarily) a lot of entry occurred from Communist Bloc lines and national lines of less developed countries. It is doubtful that these shipping lines operated on a profit-maximizing basis or even that they earned positive profits. Therefore, the modern era does not provide as clean a market for analysis.

of three conferences studied here cover routes from the UK to British colonies (U.K to South Africa, U.K. to India), the third runs from the UK to the Far East (China, Japan, Hong Kong); all are composed of mostly British firms. I include both directions of the route in my discussion and analysis; these three conferences had a high proportion of overlapping members in the two directions.

Besides fixing rates of freight and volumes carried, the conferences payed deferred rebates that protected their monopoly. To promote shipper loyalty, the lines arranged to pay back a certain percentage (usually 10%) of a shipper's total freight bill over six months, if the shipper had patronized the conference exclusively during that six months and for a further six or nine month waiting period. Therefore the shipper faced the loss of 10% of a year's freight if he decided to switch shipping conferences. The penalty could have been designed to promote stablility in the amount of cargo shipped which would lower average costs per ton and reduce risk for the shipowner. However, a bulk (or regular) shipment discount could have achieved that end. The deferred rebate had the anti-competitive effect of discouraging other shipping lines from entering.

Cheating on the part of cartel members was not a significant problem. A convenient feature of shipping from a collusive point of view is that it is impossible to expand capacity (number of ships) without being detected.[12] Secret price discounts could have been given, but as quantity was essentially fixed, the shipper would not have gained significantly. Revenue pooling also cut down on the incentive to attract cargo

[12] Some cheating in the way of undercounting tons shipped did occur in the 20s and 30s in the Far East.

away from other lines and promoted stability by giving the aggressive line only a fraction of marginal freight revenues. A revenue pool might require every line to contribute 50% of its freight revenues, for example, to a common fund. Distributions from the fund were made in accordance with previously agreed upon shares. Revenue pools, direct payments, and cargo sharing agreements existed in all conferences. Brokers in a port would have many merchants as clients and perhaps even work for more than one shipping company. These brokers provided another limit on cheating; they could leak news of the violation to other brokers or members of the cartel. Since brokers earned percentage commissions, lower rates were not in their interest.

Market Power

Whether or not the conferences were monopolies is critical in analyzing the motivation for the observed price wars. The only substitutes available to the merchant were sailing ships (risky and slower), tramp ships, or a specially chartered vessel. Tramps were a viable option in a port such as Singapore where many ships passed through on their return journey to Europe with empty space.[13] In South Africa tramps were rarer, much slower, and did not reliably have space at any given time. For all the conferences, speed was probably the most crucial distinguishing characteristic of the conference ships. Relatively valuable manufactures and tea had high opportunity costs of time. Slower means of transportation were not good substitutes.

The chartering solution would not be feasible for a merchant who wanted to send

[13] In fact, the "Straits Conference" from Singapore was very unstable and much less successful than others due to tramp competition.

regular, less-than-shipload lots; too much space, capital, and risk was involved. Additionally, the conference defended itself vigorously against charter companies; those merchants who joined together to form a charter company would be discriminated against by the conference during their rebellion and perhaps afterwards also. The final feature adding to the conference's monopoly power was the dissociation of ownership of goods and arrangement of transport. The broker profited from higher rates and was unlikely to work against the conference.[14]

There is anecdotal evidence that prices in all three conferences generated positive profits for cartel members. During price wars, the entrant line was occasionally able to operate without loss at the fighting rates. For example, Solomon reports that during the price war caused by the entry of shipowner Houston in the South African trade, Houston's half price rates were enough to cover his costs and the costs of the slower Conference lines.[15] Regulated prices kept demand low and carefully restricted sailings stopped lines from undermining the monopoly equilibrium with higher quantity. Letters from the owner-manager of the Blue Funnel line report that Far East Conference rates increased Blue Funnel earnings by 20% over free market rates.[16] The Calcutta Conference members claimed to be earning profits after cutting their prices in response to entry by India Mutual. The conference markup was substantial in all three markets,

---

[14] The Far Eastern Freight conference is known to have given a 5% kickback to a group of large and influential brokers in Singapore in order to direct more freight to the conference group in a port full of tramp shipping. Brokers booked the freight of their clients and on occasion kept the 10% rebate unless the client was experienced enough to ask for it.

[15] F. Solomon (1982).

[16] Hyde, Enterprise p96.

although probably greatest in the South African cartel.[17]

Accounting Data

To give the reader some idea of the costs and profits of steamship operation, a selection of data in nominal Pounds Sterling follow below. These data include profits, rates of return, average cost, and shipbuilding costs.[18] Further tables can be found in Appendix I.

In the 1870s steamships for international voyages were only 1,000-2,000 gross tons. By the 1880s, common new ship size was in the 3-4,000 ton range. In the nineties, big technical advances in engines and twin screw propellers led to much bigger ships of about 6,000 tons. A large steamship of 7,000 gross tons cost about £70,000 in 1893. Such a ship would have a life of approximately 20 years and would earn a net annual income of £7,000. Size continued to increase up to WWI, although different trades had different needs. For example, North Atlantic passenger ships were much bigger than cargo ships intended to travel to South Africa.

Technological advances vastly changed nearly every aspect of steamship management over the period. For example, the time needed for the voyages decreased substantially. In the early 1870s a steamer could reach South Africa in about 25 days.

---

[17] The elasticity of demand for shipping services is carefully treated in Bennathan and Walters. They discuss the full derivation where t is cost of transport, P is price, E_d and E_s are elasticity of demand and supply for the transported good respectively:

$$E_{transport} = t/P_d\{E_sE_d/(E_s - (1-t/P_d)E_d)\}$$

The Marshallian elasticity can be found by setting the elasticity of supply equal to infinity. Then the equation becomes:

$$E_{transport} = E_d*t/P_d$$

[18] The principle collector was Professor Francis Hyde of Liverpool University who worked with the records of the Harrison and Holt family firms.

By the early 20th century South Africa was only 10 days away. As a result, a firm wanting to run a weekly service to South Africa could own fewer ships than would have been necessary 25 years before - although those ships would be more expensive and bigger. Total costs per ton mile are reported in Appendix I. Average costs fell over time with technological improvements to engines and propellers and as ship size grew. Increasing returns to scale certainly existed at a single ship level, hence the continuous growth in the size of ships. Cargo handling costs were the main element of marginal costs. Kevin Burley estimates that marginal cost for Australian shipping in 1923 were about 18-19% of total costs.[19] Taking this figure to be roughly correct implies a "fighting rate" for freight could be 80% lower than average cost and still cover marginal costs.

In the 1870s the Charente Steamship Company earned £9-11,000 per voyage to India in freight, with costs remaining relatively stable at about £7,000, leaving profits of £2-4,000 per voyage.[20] The rate of return earned by Harrison ships in the period 1895-1914 was over 6%.[21] Hyde writes that OSS (Holt family firm) earnings were somewhat lower for trips to China in the 1870s and 1880s; average net profit was £2,000 per voyage.[22] Average net voyage profits for the Far East route mentioned above ranged from £620 in bad years to £4,000 in boom periods. Of course, revenue pooling

---

[19] K. Burley

[20] Hyde, Harrisons p40.

[21] Hyde, Harrisons, p119.

[22] Hyde, Blue Funnel p69.

arrangements smoothed out differences among companies and commodities, but the industry tended to be very cyclical. From 1897-99 the value of exports to South Africa fell at an average rate of 6%; the next four years it grew at 15,35,40,and 2%, respectively. Steamship companies had to expect and plan for such fluctuations. Price wars also disrupted profit flows; Harrisons earnings fell from £143,000 in 1891 to £72,000 in 1892 because of price wars with Brocklebank, Anchor Lines and Indian Mutual.[23]

Cargo rates functioned rather like modern day airline rates. Each type of good had a different rate, often nonlinear in quantity, as did each port-pair. On top of that, shippers might or might not be able to claim their deferred rebate. As a result, it is nearly impossible for a researcher to construct an average per ton rate; I do not attempt it in this paper.

IV. Conference Histories

South Africa

The South African conference served the ports of the Cape Colony and Natal from both the East and West coasts of Britain (see Figure 3). The two founding lines were the Union Line and Castle Line. In the 1870s the South African Government gave each a half share in the mail contract to spur competition; the Union and Castle lines split the

[23] Hyde, Shipping Enterprise and Management, p96.

contract until 1900 when they merged.[24] Meanwhile steamships were becoming more efficient and profitable and a spate of entry began. Detailed descriptions of each case of entry are omitted from this version of the paper. The Union and Castle lines felt sufficiently threatened to organize a formal conference in 1882. The conference established freight rates to be charged by all members and tightly regulated the number of sailings and ports of call allotted to each member line. The two mail lines worked closely in the conference, essentially deciding the rules for the other member lines; they kept the lion's share of sailings for themselves and negotiated favorable rate differentials for their faster (mail) ships. Additionally, South Africa did not export very much during the period - gold and diamonds were the major commodity exports - yet the colony demanded every sort of British manufactured good from railway ties to jam. Needless to say, the shipping lines did not have much volume in the way of "homeward," or UK bound, freight. The two mail lines expressly forbade the others from loading what little homeward freight there was, keeping this market for themselves.[25] The mail lines made substantial side payments on several occasions to settle disputes and most lines belonged to at least one revenue pooling arrangement. Though the cartel was disrupted by entry, price wars never arose because of internal disagreements alone.

Competition on the basis of price, though submerged, is occasionally visible in

---

[24] Though the South African Conference was not formally established until 1883, I include entry before that time because price wars existed then also. The periods of the Boer War (1900-02) and World War I (1914-18) are excluded from the story as regular merchant shipping services were suspended.

[25] In fact, a cargo line which "loaded homewards" was fined £2,000 by the conference! Solomon p39.

shipper actions. For example, a line could compete on the basis of price by absorbing wharfage dues, usually paid by the shipper, or by including some inland transportation. The rate differential between the faster mail ships and the slower cargo lines was also an area of continual disagreement, as each member line tried to implement a pricing policy favorable to itself.

A particular characteristic of South African trade was the strength of local merchant feeling against the monopoly power of the Conference.[26] On several occasions South African merchants who wanted an alternative to the conference encouraged entry without success. Local merchant associations occasionally formed their own line with charter ships, or threatened to, in all three conferences. If the merchants were homogeneous and united, as in the case of Calcutta, they could gain. Most often merchants movements were too fractured to extract concessions from the conference, as in South Africa. Conference pricing gave merchants one advantage, "...a uniform, continuous rate to make forward contracts with a certainty that no competitor by getting cheaper conveyance, can undercut him or depreciate his stock."[27] Merchants could earn windfall profits and losses if freight were set by a free market because freight composed a large part of the cost of a merchant's goods and changed by substantial amounts without conference regulations.

Calcutta and Bombay Conferences

The Indian conferences were naturally more disparate than the South African

---

[26] See Solomon's text, South Africa., for a thorough description.

[27] Hyde, Shipping Enterprise, p97

conference due to the geography of the route (see Figure 1). Some members continued on to the Far East, some to the Persian Gulf, some to the East coast of Africa, etc. The first Calcutta conference was formed in 1875 consisting of British India, the P&O, Hall, and City lines.[28] All Indian tea was shipped from Calcutta or Ceylon; the Calcutta Conference agreed on the tea rate each year. Each line paid a percentage of its freight rates into a common pool which was then allotted to lines in a certain proportion. No restrictions were placed on volume carried or number of sailings. Again, many side payments and other pooling arrangements also existed. For example, after Clan entered the tea trade in 1882, it was paid £2,000 per year by the P&O and £1,000 by British India not to carry passengers.[29]

On occasion, "entry" (into the competitive fringe) would occur because a member quit the conference over some dispute about its share. Only seven entries in the sample are this type; a representative case occurred in 1886 when Clan decided it would like a bigger share of the tea coming from Calcutta. The other lines were not receptive, so Clan withdrew from the Conference and started carrying tea in special arrangements with plantations. The conference capitulated in the summer of 1887, giving Clan a larger percentage of trade. Motivations for such behavior were different from standard entry and are treated separately in the estimation.

The Bombay Conference was initially formed in 1879. It set standard cargo and

---

[28] The P&O was the longest established major carrier, being heavily subsidized by the UK Government. Early on it was concerned about its monopoly position in the carriage of opium, an extremely lucrative trade, and made several bilateral deals before the conference was formed.

[29] Muir p133.

passenge rates, most importantly rates on Lancashire cotton piece goods from Liverpool. By 1885, its members consisted of the P&O, British India, Anchor, Harrison, Hall, Clan, City, and Rathbone.[30] I can find 16 cases of entry occurring in the following years, much as in the South African trade. Many entrants were resented and fought vigorously, others were given complex shared rights to certain ports and cargo.

Throughout the study, the continental ports of Hamburg and Antwerp are considered to be good substitutes for London, Liverpool, Glasgow, and Middlesbrough. A map of the area (Figure 2) shows how geographically close these ports are to each other, particularly in comparison to the long trade routes considered here. Additionally, Germany produced many of the same manufactured goods in demand in the Far East and South Africa, so transshipment of cargo was not necessarily required. German lines did enter all three cartels during the period and provided robust competition for the British lines.[31]

Far Eastern Freight Conference

The original Far Eastern Freight Conference (hereafter FEFC) agreement included the P&O, OSS, Glen, Shire, Mogul, Skinner, and Messageries Maritimes (French) in 1879. The most valuable cargo of all heading to China was Lancashire and Yorkshire (L&Y) goods: yarn, wool, cotton, and silk manufactures. OSS, being based in Liverpool, had a virtual monopoly of L&Y goods. However, it paid shares of the L&Y

---

[30] All were original members except for Rathbone. Taylor, p213.

[31] A large literature exists on the topics of growth of German shipping, its competition with British shipping, and the role of subsidies during the period. See D.H. Aldcroft ed., Studies in British Transport History 1870-1914, 1974, his own article entitled "British Shipping and Foreign Competition: The Anglo-German Rivalry, 1880-1914."

revenue to other lines in the Conference which were forbidden to load at Liverpool.

Although familiar names tried to enter in the early period of the FEFC, in the 1890s a new group of non-British lines began sending ships to China. A few nations which had interests or possessions in the region now had ocean-going steamships; Swedish and Italian lines are the main examples.[32] Austrian Lloyd (Italian) entered Far Eastern routes in 1892; it was heavily subsidized by the Italian Government, but seemingly without a clear strategy in mind for the business of the line. In contrast, the government of Japan gave the FEFC's main competitor, the newly formed Nippon Yusen Kaisha (NYK) line, explicit goals. One was to keep export freight rates low; NYK frequently was denied rate increases on silk and cheap manufactures. The NYK bargained and muscled its way into the conference in several stages from 1893 to 1901. Government subsidies, special route subsidies, and specific rate requests on the part of the government considerably enhanced NYK's bargaining power *vis a vis* the FEFC.[33]

V. Data and Estimation

The data I use were collected from firm histories of shipping companies. Other books listed in the bibliography are general histories of shipping or of a specific port. My dataset includes when, where, and for how long price wars occurred; these facts were reported in the texts just mentioned. Four potential observations have been dropped

---

[32] I have not been able to find more than one passing references to the Swedish line, so it has been excluded hereafter.

[33] Doug Irwin (1991 JPE) applies modern strategic trade theories to this period. He does not claim the agents knew what they were doing strategically, whereas the in this case the Japanese Government may well have been aware of the effects of its actions.

due to insufficient information on tonnage, subsidy, or exact dates of events. Table II holds summary statistics for the different types of entrant while the schematic table below shows the number of observations in different categories. "Dispute" refers to price wars begun by defecting cartel members.

	South Africa	India	Far East	Total
Entries	24	13	15	52
Price Wars	7	7	6	20
Disputes	0	3	5	8

Costs of a Price War

Clearly the shipping conferences felt that keeping their monopoly positions was worth periodic price wars. However, they did not fight every entrant which threatened to erode the conference's monopoly position. A price war is an expensive way to protect a monopoly. The first concern of any type of entrant or cartel would be the cost of fighting a price war. All three hypotheses above claim the monopolist gains from predatory pricing; before undertaking a price war, the monopolists needs to weigh those benefits against the costs. It is impossible to determine the expected cost of a war without detailed information on marginal cost for each ship in each line, composition of cargoes, exact structure of rate cuts, etc..[34] Not only that, a conference had the power

[34] If the entrant were serving a selected port or carrying a particular commodity, the price cuts could be specifically targeted and the cost to the conference members much reduced. The war against DADG's entry to a single South African port in 1894 is a perfect example; the conference lowered rates on fertilizer and machinery to that port. I cannot find this information for each observation.

to choose the magnitude of the price cuts if it began a war.

However, I can identify some variables which track the costs of fighting a war. Regardless of whether the entrant has bought or leased its ships, it loses the opportunity to use its vessels on another route that period if it stays in the war. Unlike many other industries, capital expenditure is not really a credible commitment to stay on a route since ships are fungible. Unless opportunities on other routes are much worse, ships have a similar value elsewhere; entrants cannot sink large investments to raise the cost of predatory pricing and discourage it. The entrant's opportunity cost, which here is trade growth on other routes, will affect incentives to enter and exit. *Trade Opportunities* is defined to be the growth rate of all British trade excluding the entrant's route in real Pounds Sterling.

Trade Growth is a two-year moving average of the percent change in the value of trade on the route in real Pounds Sterling.[35] The value of *Trade Growth* fluctuates widely from year to year because of extremely procyclical trade flows during the late 1800s. If trade on a route was increasing the conference may have been less likely to fight a price war; the entrant's additional capacity was not causing as much absolute profit loss as it would during a slump and a price war would mean more foregone profits than a war during a recession.

Route Tons is the number of gross tons the entrant places on the route measured

[35] Prices were reasonably stable over the period. Inflation rates did not rise above 6% during the period.

in thousands of tons.[36] The amount of tonnage the entrant places on the route determines how much profit the cartel loses. The conference should also care about the relative size of the entrant's route tonnage (*Size = Route Tons/Cartel Tons*). As the size of the successful entrant rises, either *ex post* Cournot[37] prices (including the new entrant's capacity) will fall or else current members' share of perfect monopoly profits will fall. Greater lost profits from large entrants lowers the opportunity cost of any type of price war. If any of these cost variables affect the likelihood of a price war then we can conclude that the conference is acting, at least partly, on the basis of a profit-maximizing motive.[38] Below, the three theories of strategic price wars are discussed in conjunction with variables which might affect the prevalence of one or the other. Several observable characteristics affect the cost of predation as well as provide motivation for one or more strategic theories. I will identify which variables can distinguish specific theories.

Long Purse

To test the long purse hypothesis one ideally requires data on the financial

---

[36] I realize that net tons would be a better measure of capacity than gross tons (Net tons measures only cargo space whereas gross tons includes the space taken up by the engine and fuel). However, gross tonnage data were available for all observations while net tonnage data were not.

[37] Cournot may be a reasonable assumption for this market; the cartels seem to have set capacity in preparation for a volatile and cyclical market, and then chosen prices that kept some of that capacity empty in non-peak periods.

[38] An alternative explanation for price wars could be irrational behavior on the part of either incumbent or entrant owners. The period being analyzed is 1875-1929 when capitalism was not highly developed; it could well be the case that friendships, connections, and pride were stronger motivations for shipowners than costs and profits. If the cost variables have the expected signs then we can conclude shipowners wanted to maximize net profits.

backing of an entrant: personal resources of the main owner and his family, resources of other equity holders, relationship with bank, and outstanding loans. Much of this information is unobservable to me. However, I do know whether the company was publicly traded on the London Stock Exchange or not and can use that as an explanatory variable (*Publicly Owned*).[39] The long purse hypothesis is less likely to be a motivation when the entrant's financial situation is publicly known; a public firm can conceal less information than a privately owned firm. Additionally, trying to make a government subsidized line (*Government Subsidy*) run out of funds and exit does not make much sense since the firm has a soft budget constraint.

Firm Tons is the total number of thousands of gross tons that the entrant owns; all of them may not be on the route in question. The size of the entrant firm is correlated with its financial and intangible resources. A very large firm would be less likely to provoke a price war after entering a given route. The very biggest lines had large cash reserves and networks of ships all over the world. The dummy variable *Big* takes a one if the firm's total tonnage is greater than 125,000 gross tons. The ability to self-insure and spread risk, information about trading opportunities, and frequent design and purchase of ships made entrants from *Big* firms stronger. All the tonnage variables are proxies for accumulated resources (*Route Tons, Firm Tons*).

The dummy variable *New* is assigned a one if the firm had existed for five years or less. A young firm is unlikely to have the cash and insurance reserves common to

---

[39] Stock Exchange Intelligence lists publicly traded shipping companies' financial structure in its yearly report. Most of my entrants were not publicly traded; I have the debt and equity information for those that were which was certainly common knowledge.

established shipping firms at the time. After entry, the conference would likely think its own reserves were greater than those of the new firm; a price war could have begun for long purse reasons. Contracts (*Contract*) secured by an entrant before entry will make a long purse price war less effective, therefore longer and more costly to the conference and in turn, less likely to be undertaken.

Renegotiation or Bargaining

As the entrant's tonnage on the route (*Route Tons*) increases, it is likely to provoke price wars for renegotiation reasons; the cartel would like to drive down the entrant's route tonnage even if it does not expect to drive it to zero. The dummy variable *Dispute* receives a one when a dispute within the cartel caused a member(s) to quit and then re-enter. In my dataset, all disputes involved price wars before being settled; that is how they are defined. The defecting line or lines generally wanted a larger share of cargo or profits. A member who quit knew the reputation, wealth, cash flow, and internal politics of the cartel, so reputation and long purse are unlikely motivations for the war. However, the cartel still had to show the defecting member to what extent it could be "held up," so signalling was a major motivation.

A line receiving a *Government Subsidy* could not be driven into bankruptcy in the same way as private lines. However, a war against a subsidized line could be effective in convincing it to reduce its services on the route.[40] Again, long term contracts may

[40] Non-monetary as well as monetary penalties existed for managers whose firms earned persistent operating deficits. Both the German and Japanese subsidized lines experienced problems gathering and keeping political support for their subsidies, especially when they made losses. NYK had an unstable political constituency in the Diet in the 1880s; its subsidies were approved with difficulty and the debate affected its status in the conference. NDL lost 5.25 million Marks on mail steamer lines to East Asia and Australia despite a subsidy of 44.3 million

have reduced the likelihood of a signalling price war; the entrant had to maintain sufficient tonnage to fulfill the contract and that tonnage would be immune to a price war. A trade boom (*Trade Growth*) reduces the need to bargain down the entrant's quantity in absolute terms. Finally, outside trade opportunities (*Trade Opportunities*) may have altered the incentives for a renegotiating-type price war. If other routes offer unattractive opportunities, the cartel would have to wage a more vigorous price war in order to force exit of some entrant capacity.

Reputation

It is more difficult to test reputation models since their implications are sensitive to the assumptions of the model. I explore reputation-motivated wars somewhat by defining and including two dynamic variables. *Time Since Last Entry* measures the number of years since the last entrant entered that cartel, while *Time Since Last War* measures the number of years since the last war occurred in the cartel. (Neither of these two variables is defined for the first entrants in each cartel.) If the cartel is signalling that it is a tough type with each price war and its reputation erodes, there will be less need for a reputation-building war if one just occurred. Therefore, the probability of war conditional on entry will increase as the last war recedes in time. However, if the cartel's reputation does not erode and is well-known before the sample period starts, this dynamic effect will not show up in the results. *Time Since Last War* and *Time Since Last Entry* also have cash flow implications. A firm which recently experienced low cash flow because of an war or entry might find it more costly to undertake another price war.

Marks. Scholl in Yui, p200.

Past War Percentage is the cartel-specific percentage of entries which have resulted in wars up to the current entry. *Past War Percentage* may be the equilibrium percentage of wars given a certain repuation for that cartel.

Table II shows the expected effect on the probability of war for each of these variables under each strategic hypothesis. If the variable directly affects the costs of a price war, regardless of the type of entrant, the influence on the likelihood of a war is noted in the last column. Long purse motivations can be seen in the variab · *Publicly Owned, Contract, Firm Tons, Big* and *New.* If renegotiation is a cause for wars, its influence should show up in the coefficient of *Trade Opportunities, Trade Growth, Route Tons, Government Subsidy,* and *Dispute.* Reputation is only tested by the inclusion of *Past War Percentage* and *Time Since Last War.*

Estimation

The data I have were common knowledge at the time of entry: tonnage, age of the firm, trade figures. These characteristics are only a component of the true type of the entrant. The variables the econometrician observes were also components of the signal of the entrant's type which the cartel received. In a sense, I have a signal of entrant type which is worse than the cartel's signal. When the cartel receives a high signal, it does not know if the true type of the entrant is high, or if the signal is high and the true type low. In either case it will share the market; wars will be negatively correlated with the cartel's signal. However, the cartel's signal is correlated with the data I have. Thus, regressing a vector of common knowledge characteristics on whether

or not a price war occurred will give an estimate of which characteristics are important in causing price wars.

$$\text{Prob}(\text{ War } | \text{ Entry }) = \Phi(\, X\beta \,)$$

where the dependent variable is zero in case of an uncontested entry, one in case of a price war. The X's are the characteristics of the entrant and the market, most of which have been discussed previously,

$$\begin{aligned} WAR = \; & \Phi(\beta_1 RouteTons + \beta_2 RouteTons^2 + \beta_3 FirmTons \\ & + \beta_4 FirmTons^2 + \beta_5 New + \beta_6 Big + \beta_7 TradeGrowth \\ & + \beta_8 TradeOpportunities + \beta_9 Contract) \end{aligned}$$

A second interesting question is what size, or route tonnage, the entrant decides to put on the route, conditional on entry. One would suspect that previous play of the game as well as entrant characteristics will affect the entrant's choice of *Route Tons*. Modelling such an interaction is beyond the scope of this paper, but I present a few empirical results with an informative motive.

$$\text{E}(\text{ Route Tons } | \text{ Entry }) = X\gamma$$

Past War Percentage may also help explain choice of route tons. The percentage of

price wars in the past may be correlated with the probability of one occurring in the present. Conditional on entry, a line will want to place more tonnage on a route if it is about to be bargained down in a price war.

Results

Results from the probit estimation are contained in Table IV. I use an adjusted tonnage variable, *Normalized Route Tons*, which is formed by weighting *Route Tons* by the inverse of the route's distance.[41] Figure 4 shows the combined effect of the *Normalized Route Tons* variables. An entrant putting from 40,000 to 70,000 tons on a route faces the least chance of a price war, all else equal. Smaller scales of entry attract price wars with greater probability. However, it is the really large tonnage entrants which see the biggest increase in the probability of a price war. A larger number of *Route Tons* does more damage to cartel profits; the opportunity cost to the cartel of starting a price war falls and the opportunities for renegotiation rise. Thus, the cartel is more likely to start a renegotiation price war to reduce the entrant's tonnage. The effects of *Firm Tons* and *Big* are plotted in Figure 5. The dummy variable *Big* has a large negative coefficient; the entry of very big lines is not often met with an aggressive response. A long purse explanation for price wars is consistent with this finding. The largest overall effect of the tonnage variables is again for the biggest firms; they are much less likely to be fought. The very tiniest firms appear to be less often fought than slightly larger, but still small, firms. The coefficients on both *Route Tons* and *Firm Tons*

[41] Indian observations remain unchanged.

suggest that cartels may not have wasted resources bothering with tiny firms. *Size* turns out to be a much weaker explanatory variable than *Normalized Route Tons*, so it is not discussed further.

New has a large positive coefficient; young firms are more likely to be fought by a cartel. A long purse theory could explain this result; *New* firms are less likely to have the accumulated resources that an older firm has. However, the cost of the war may also be correlated with the age of the entrant, if age is associated with characteristics such as less experienced brokers. Both of these effects are showing up in one coefficient. The extent of *Trade Growth* on the route is critical in determining the probability of a price war. If trade is growing quickly, more profits are foregone in the event of war and the cartel is less likely to initiate one. Renegotiation of the size of the entrant is more difficult to undertake in the midst of a boom. Positive outside *Trade Opportunities* increase the probability of a price war. The entrant could be employing its capital elsewhere in profitable way and so the cartel's expected cost of a renegotiation price war is less. Having a contract to carry cargo on the route (*Contract*) reduces the probability of a price war; the cartel realizes a war to persuade the entrant to reduce its tonnage would be costly and ineffective. Surprisingly, the conference dummies are always insignificant and are therefore omitted.

The lines which started price wars because of a dispute within the cartel are likely to have had somewhat different motivations and characteristics. It may not be appropriate to combine the two groups in the estimation. Therefore, the second and third columns of Table IV omit "dispute" observations. These specifications show bigger

coefficients as well as a larger loglikelihood value. *Normalized Route Tons*, *New*, and *Contract* become much more significant. *Publicly Owned* and *Government Subsidy* do not affect the probability of a price war and are not included in any specification. They are probably just too crude to capture any information not already incorporated in other variables. Similarly, neither a *Via the Cape* line nor a *Branch* line alters the chance of a war. There does not appear to be a significant trend in the probability of a price war over time; *Time* is insignificant if included.

In the third specification I add the variable *Past War Percentage*. The cartel's past behavior may have created a reputation to which entrants are reacting. The coefficient is positive and marginally significant; a higher equilibrium-war-percentage cartel is more likely to fight the next entrant. This result gives no support to any dynamic story of learning or erosion of cartel reputation. Instead, we might think that some cartels have a higher equilibrium percentage of wars and this variable is picking up inter-cartel variation in frequency of wars. Neither *Time Since Last War* nor *Time Since Last Entry* are at all significant if they are included in this specification. Table V gives the marginal effects for the results in Table IV.

Table IV's coefficients are almost all significant at the 10% level despite the small number of observations. The explanatory variables, characteristics of entrants and the market, are empirically important in predicting whether or not a cartel will begin a price war. The significance of those characteristics provides evidence for a predatory pricing explanation of price wars, rather than a demand or cost shock approach.

The second estimated equation is an OLS regression explaining an entrant's route

ton choice conditional on entry. The results are reported in Table VI. The total amount of tonnage the firm owns (*Firm Tons*) has a small but significant influence in predicting how much it will enter into the new route. In column one *Dispute* has a positive coefficient; as expected, an established member of a cartel is likely to be bigger than an average entrant.

The trade variables are very significant and of the expected sign. Increased trade on the route (*Trade Growth*) leads to entry with more tons; relatively good trade elsewhere (*Trade Opportunities*) causes an entrant to place fewer tons on the route. Given that an entrant has decided to enter the route, it will want to enter with more tons if there is a history of price wars on the route. Price wars serve as renegotiating tools; if an entrant is likely to face a price war, it's better off having more tonnage to bargain with and more tonnage to fight with. *Past War Percentage* has a positive and significant coefficient. The conference dummy variables should have some power in the regression because I am not adjusting route tons for distance. Firms enter with more tonnage on the Far East route, which is logical because it is the longest. *New* has a large negative coefficient, but is not very significant; young lines enter with a smaller amount of tonnage. *Contract* behaves much the same way; it is not really significant but of plausible sign and magnitude. *Government Subsidy* is insignificant if included in this specification.

In column two I limit the regression to observations which were not disputes. The results are very similar except *New* and *Contract* increase in significance. *Past War Percentage* declines in both significance and magnitude in this specification. *Dispute*

observations must be driving some of the effect of *Past War Percentage*. Firms which quit their cartel are expecting to renegotiate the size of their fleet on the route. If past wars are correlated with current wars then the firm should begin the negotiation with more tonnage as *Past War Percentage* rises. The final column of Table VI includes only observations where *Time since Last War* is defined. Recall that the first observations in each cartel have undefined *Time since Last War*. The coefficient is insignificant (partly because the number of observations drops). If a cartel's reputation is well-known, then the amount of time since the last war should not affect the entrant's choice of *Route Tons*. This result is therefore not inconsistent with reputation stories, but is not strong evidence in their favor.

VI. Correction for Misclassification of a Price War

It is possible that my sources may mistake a negative demand shock for a predatory price war. The reason this is of some concern is that if the dependent variable in a probit regression is misclassified, the resulting coefficients are biased (Hausman and Scott Morton (1994).[42] Misclassification could occur in the other direction also: an entry which resulted in a price war, but was never described as such in my sources. If a price war is misclassified, unnoticed or wrongfully described as a war, then the estimates presented in Table IV will be inconsistent. Hausman and Scott Morton (1994) discuss two methods for correcting such bias. One is simply to use a likelihood function

[42] Porter (1983) uses econometric techniques to identify railroad price wars and discovers a high correlation between his results and accounts of price wars cited by newspapers. His evidence suggests historical reports of price wars were quite accurate.

which explicitly includes the possibility of misclassification which we call α. Maximizing the likelihood function results in an estimate of the percentage of observations which are misclassified as well as estimates of the other coefficients. The second method is semiparametric and does not depend on an assumed error distribution. The semiparametric method uses two techniques. First we use Han's (1987) Maximum Rank Correlation (MRC) procedure to estimate the coefficients of the explanatory variables. Secondly, we estimate the extent of misclassification with the Isotonic Regression (IR).

Table VII repeats the earlier probit results and gives both the MLE and MRC/IC coefficients. All of these methods estimates the ratio of the coefficients rather than their absolute size. The MRC procedure estimates a normalized coefficient vector which must be rescaled by holding one variable fixed. (*Contract* is held fixed in Table VII.) MLE results which are robust to misclassification, but assume a normal error distribution, are in column two, and semiparametric results which are robust to both misclassification and non-normal error distribution are in column three. Column two indicates that some entries are misclassified as "no war" when they should be "war." The point estimate for the extent of "no war" misclassification, α_0, is 9%, but the standard error is 6%. The estimate of α_1, "war" misclassification, is almost zero; that result indicates that observed price wars are essentially never the result of misclassification. The standard errors rise across the board when the new coefficients are included in the estimation. All coefficients change at least slightly between the standard probit and MLE results. The coefficients on the trade variables, *Big*, and *New* are now larger, while route tonnage

coefficients decline dramatically and become insignificant. Except for the coefficients on the Route Tons variables, the same general conclusions can be drawn from the MLE results as the probit results.

However, the MLE results rely on the assumption of normally distributed errors, unlike the semiparametric results. The semiparametric coefficients lie for the most part near or below the original probit estimates. Both *Route Tons* coefficients return to their former magnitude. Isotonic Regression method estimates the first step at height zero and the last at height one; the estimates of both α_0 and α_1 are zero. These estimates tell us that neither category of dependent variable suffers from misclassification.

Because the estimates are close to those of the probit specification assuming no misclassification, the results suggest that the assumption of normally distributed errors is a good approximation for these data. Since all the probit restrictions appear to hold, in theory the three columns of results should be the same. In column two, Maximum likelihood with 44 observations does not estimate the coefficients as accurately as the MRC/IR method, despite normally distributed errors and the absence of misclassification.[43] The semiparametric method is a superior technique. However, the standard errors on the semiparametric estimates are very sensitive to the window width chosen for their estimation. There is little theoretical guidance on which window width to choose, so I report two possibilities, although the standard errors differ dramatically

[43] Simulations in Hausman and Scott Morton (1993) confirm this point.

between the two.[44]

VII. Conclusions

The results of the analysis here suggest that British shipping conferences could have employed predatory price wars against entrants for renegotiation and long purse reasons. Opportunities for outside trade, growth of trade on the route, the age of the entrant, the size of the entrant's firm, and the size of the entrant on the route all contributed strongly to the cartel's decision to prey. The coefficients provide support for the many theoretical papers which conclude predation will be a more profitable strategy against "weaker" entrants. The characteristics which I find make an entrant weak are lack of age and the experience and financial resources correlated with age. Higher absolute size of the firm, regardless of its entry size, makes the entrant stronger. Additionally, an entrant with a contract to carry cargo on the new route is much stronger than an entrant without such a contract. Government subsidies and public ownership do not affect the entrant's strength. These results confirm long purse and renegotiation theories of predation. I have little evidence to support any kind of reputation motivation for price wars. The only variable I have which might support a reputation story also measures the cost of a price war and the two effects work in opposite directions.

Entries that were really "breaks" by existing members of the cartel are not explained as well by my specification. I conclude that the motivations for these

---

[44] When we calculate standard errors for the coefficients in Hausman and Scott Morton (1993), they do not vary much at all with window width. There, we use a sample of job-changers from the CPS that has a total of about 5,000 observations.

observations were quite different from those of standard entrants. The results also refute the theory that price wars in this industry were caused only by competition (demand or cost shocks) rather than predatory behavior. If the wars occurred simply because of market conditions, then the variable *New* should not be significant. The age of an entrant line should be irrelevant to the opportunity cost of entering or exiting the route, the size of any shock, or the cartel's cost of a price war. In fact, *New* is significant in predicting whether or not a price war will be undertaken against an entrant, suggesting long purse predatory motives on the part of the cartel. Not only market characteristics such as trade, but entrant specific characteristics (other than quantity) matter in predicting price wars. Thus, I find empirical support for theoretical models that allow the likelihood of war to depend critically on the type of entrant.

References

Aldcroft, Derek H. (1968), "The Mercantile Marine" in The Development of British Industry and Foreign Competition 1875-1914, London.

Anchor Line Ltd. (1932), The Book of the Anchor Line, ED.J. Burrow & Co., London. (Baker WNP/A539/A53a)

History of the Anchor Line(1911) printed by John Horn Ltd., Glasgow. (Baker WNP/A539/A53)

Bennathan, Esra, and Walters, Alan (1969), The Economics of Ocean Freight Rates, Praeger, New York.

Benoit, Jean-Pierre (1983), "Entry with Exit: An Extensive Form Treatment of Predation with Financial Constraints," IMSSS Technical Report, No 405, Stanford University.

Benoit, Jean-Pierre (1984), "Financially Constrained Entry in a Game with Incomplete Information," *RAND Journal of Economics*:15:4:490-499.

Blake, George (1956), The Ben Line, Thomas Nelson and Sons Ltd.. (Baker WNP/T486/B36)

---- , (1962) Gellatly's 1862-1962: A Short History of the Firm, Blackie and Son Ltd., London.

Bolton, Patrick, and Scharfstein, David (1990) "A Theory of Predation Based on Agency Problems in Financial Contracting," *American Economic Review*:80:1:93-106.

Bork, Robert, (1978), The Antitrust Paradox, Basic Books, New York.

Bowen, Frank C. (1939), The Flag of the Southern Cross, Charles Birchall & Sons, Ltd., Liverpool. (Baker WNP S534/B78)

Brackmann, Karl (1935), Funfzig Jahre deutscher Afrikaschiffahrt. Dietrich Reimer, Berlin. (Econ 4302.8)

Bromley, Raymond Curtis (1987), "Cartels as Alternatives to Contracts: The Case of Ocean Shipping Conferences." Dissertation, Economics Department, UCLA.

Burley, Kevin (1968), British Shipping and Australia 1920-1939, Cambridge University Press, Cambridge. (Baker TD/319.26/B96)

Cafruny, Alan W. (1987), Ruling the Waves, University of California Press, Berkeley. (ILS/HE/735.c34)

Carlton, D. and Perloff, (1990) Modern Industrial Organization.

Cave, Jonathan and Stephen W. Salant (1987), "Cartels that Vote: Agricultural Marketing Boards and Induced Voting Behavior" in Public Economics, Elizabeth Bailey, ed..

Chandler, George (1960) Liverpool Shipping, a short history. Phoenix House, London. (Econ 4306.260)

Cosslett, Stephen R., (1983) "Distribution-Free Maximum Likelihood Estimator of the Binary Choice Model," *Econometrica*:51:3:765-782.

Cramton, Peter C. and Thomas R. Palfrey (1990), "Cartel Enforcement with Uncertainty about Costs." *International Economic Review*:31:1.

Davies, P. N. (1969), "The African Steam Ship Company," in Liverpool and Merseyside, ed. J.R. Harris. Frank Cass and Co Ltd, London. (Econ 6075.258)

------ (1985), "British Shipping and World Trade: Rise and Decline, 1820-1939" in Business History of Shipping: Strategy and Structure, Tsunehiko Yui and Keiichiro Nakagawa eds., University of Tokyo Press, Tokyo. (HE/571/.I67)

Deakin, B.M. (1973), Shipping Conferences: a study of their origins, development and economic practices. Cambridge University Press, Cambridge. (EconP/13/70/5(37))

Falkus, Malcolm (1990), The Blue Funnel Legend. Macmillan, London. (Guildhall S 387/5)

Fudenberg, Drew, and David Levine (1989), "Reputation and Equilibrium Selection in Games with a Patient Player," *Econometrica* 57:4:759-778.

Fudenberg, Drew, and Jean Tirole (1983), "Capital as a Commitment: Strategic Investment to Deter Mobility," *Journal of Economic Theory* vol 31 No 2.

Fudenberg, Drew, and Jean Tirole (1985), "Predation without Reputation," Working Paper 377, MIT.

Fudenberg, Drew, and Jean Tirole (1986), "A Signal-Jamming Theory of Predation" *RAND Journal of Economics* 17:366-376.

Gibbs, Cdr. C.R. Vernon (1963), British Passenger Liners of the Five Oceans, Putnam and Co., London. (Baker WNQ/G442)

Gibson, John Frederic (1953), Brocklebanks 1770-1950 vol I, Henry Young & Sons, Ltd, Liverpool. (Baker WNP/B864/G44)

Green, Edwin (1985), "Very Private Enterprise: Ownership and Finance in British Shipping, 1825-1940." In Tsunehiko Yui and Keiichiro Nakagawa eds. Business History of Shipping: Strategy and Structure, University of Tokyo Press, Tokyo.(HE/571/.I67)

Green, Edwin, and Michael Moss (1982), A Business of National Importance: The Royal Mail Shipping Group 1902-1937. Methuen and Co, London. (Baker HE824/.G84)

Han, Aaron, (1987), "Non-Parametric Analysis of a Generalized Regression Model: The Maximum Rank Correlation Estimator," *Journal of Econometrics*:35:305-316.

Hausman, Jerry, and Scott Morton, Fiona (1994) "Misclassification of a Dependent Variable in a Discrete Response Setting," MIT mimeo.

Herman, Amos (1983), Shipping Conferences, Kluwer Law and Taxation Publishers, Deventer, Netherlands. (ILS/k4150.H47x)

Hieke, Ernst (1968), Rob. M. Sloman Jr. Verlag Hanseatischer Merkur, Hamburg. (Baker WNP/S634/H63)

Hope, Ronald (1990), A New History of British Shipping, John Murray, London. (WID-LC HE/823.H67)

Hyde, Francis E. (1957), Blue Funnel, Liverpool University Press. (Baker WNP/H75/H99)

------, "British Shipping Companies and East and South-East Asia, 1860-1939," in The Economic Development of South-East Asia

------ (1971), Liverpool and the Mersey: The Development of a Port 1700-1970, David and Charles, Newton Abbot. (Econ 4306/271.5)

------ (1967), Shipping Enterprise and Management 1830-1939, Liverpool University Press. (Baker WNP/H319/H99)

Jones, Stephanie (1989), Trade and Shipping: Lord Inchcape 1852-1932. Manchester University Press, Manchester. B/M 153

Juda, Lawrence (1983), The UNCTAD Liner Code: US Maritime Policy at a Crossroads, Westview Press, Boulder. (ILS/K4212.A4/1983/J8)

Kirkaldy, A.W. (1919), British Shipping: Its History, Organisation, and Importance. Kegan Paul, Trench, Trubner & Co, London. (Econ 4306/10.5/B)

Le Fleming, H.M. (1961), Ships of the Blue Funnel Line, Adlard Coles Ltd, Southampton. (Baker WNP/H75/L43)

Leubuscher, Charlotte (1963), The West Africa Shipping Trade 1909-1950. A.W. Sythoff Leyden, Netherlands. (Econ 4302.20)

MacKenzie, Charles H.C.. Michael E. Power and Ted L McDorman (1985), Liner Shipping Conferences, an annotated Bibliography, LexingtonBooks, DC Heath and Co, Lexington, Massachusetts. (Baker HE/564/A1.m32/1985)

Marriner, S., and F. Hyde (1967), The Senior John Samuel Swire 1825-98. Management in Far Eastern Shipping Trades. Liverpool University Press, Liverpool. (Econ 7138.140)

Marx, Daniel (1953), International Shipping Cartels. Princeton University Press. (Hilles/387.5/m39)

Mathais, P and A.W.H. Pearsall eds. (1971), Shipping: A Survey of Historical Records. David and Charles, Newton Abbot. (Econ 4306.271.10)

McGee J. (1980), "Predatory Pricing Revisited," Journal of Law and Economics:23:289-330.

McLellan, Roderick S. (1956), Anchor Line 1856-1956, published by Anchor Line Ltd, Glasgow. (Baker WNP/A539/M16)

Meeker, Royal (1905), "Shipping Subsidies," reprint from *Political Science Quarterly*:20:4. (Econ 4380.3.4)

Middlemiss, Norman L. (1988), The Clan Line Steamers. Sheild Publications, Newcastle upon Tyne. 387/5

Milgrom, Paul and Roberts, (1982), "Predation, Reputation, and Entry Deterrence," *Journal of Economic Theory* 27:280-312.

Milne, T.E.. "British Shipping in the Nineteenth Century: A Study of the Ben Line Papers." Chapter 13 of Studies in Scottish Business History.

Moore, K.A. (1981), The Early History of Freight Conferences, National Maritime Museum, Maritime Monographs and Reports No 51, Greenwich, London.

Murray, Marischal (1933), Ships and South Africa, Oxford University Press, London. (Econ/4302.5i)

------- (1953), Union-Castle Chronicle 1853-1953, Longmans, Green and Co., London.

Muir, Augustus and Mair Davies (1978), A Victorian Shipowner: A Portait of Sir Charles Cayzer, Baronet of Gartmore. Cayzer, Irvine, and Co, Ltd.. (HE/569.C38/M84)

Porter, Andrew (1986), Victorian Shipping, Business and Imperial Policy. Royal Historical Society, The Boydell Press, Woodbridge. (HE945.U25/P67)

Pirrong, Stephen Craig. (1992) "An Application of Conre Theory to the Analysis of Ocean Shipping Markets," *Journal of Law and Economics* 35:89-131.

Porter, Robert (1983), "A Study of Cartel Stability: The Joint Economic Committee, 1880-1886." *Bell Journal of Economics* 14:301-14.

Rickmers 1834-1959 (1960), published by Rickmers Rhederei Aktiengesellschaft, Hamburg. (Baker WNP/R539)

Royal Commission on Shipping Rings (1907), Minutes of Evidence and Appendices.

Selten, Reinhard (1978), "The Chain Store Paradox," *Theory and Decision*:9:127-159.

Sjostrom, William (1993) "Antitrust Immunity for Shipping Conferences: An Empty Core Approach," *Antitrust Bulletin*:38:2:419-423.

-------- (1992), "Price Discrimination by Shipping Conferences," *Logistics and Transportation Review*:28:2:207-216.

-------- (1989), "Price Discrimination by Shipping Conferences," *Journal of Political Economy*.

--------- (1988), "Monopoly Exclusion of Lower Cost Entry," *Journal of Transport Economics and Policy*.

Slade, Margaret E. (1990), "Strategic Pricing Models and Interpretation of Price-War data." *European Economic Review* 34:524-537.

------ (1989), "Price Wars in Price-setting Supergames." *Economica*:56:295-310.

Sletmo, G.K. and E.W. Williams (1981), Liner Conferences in the Container Age, Macmillan Publishing Co, New York. (Baker HE/7451.s515)

Smith, J Russel (1906), "Ocean Freight Rates and their Control by Line Carriers," *Journal of Political Economy*.

Solomon, Vivian Edgar (1982), The South African Shipping Question 1886-1914. Historical Publication Society, Capetown. (He903.4.z5/S64)

The Stock Exchange Official Intelligence vols. 1-45,(1882-1927) edited by the Secretary of the Share and Loan Department, London Stock Exchange. Spottiswoode, Ballantyne & Co., London. (Baker JUBD/S863)

Suslow, Valerie. "Cartel Contract Duration: Empirical Evidence from International Cartels" forthcoming *Rand Journal of Economics*.

Taylor, James (1976), Ellerman's: A Wealth of Shipping. Wilton House Gentry, London. (Guildhall S 387/5)

United States, Reports from the Consuls of (1890), "Steam-ship Subsidies" No. 112. Bureau of Labor Statistics, Washington D.C.. (Econ 4380.1)

Wray, William D. (1984), Mitsubishi and the N.Y.K., 1870-1914: Business Strategy in the Japanese Shipping Industry. Harvard University Press, Cambridge. (HE945.N72/W7)

Wray, William D. (1985), "NYK and the Commerical Diplomacy of the Far Eastern Freight Conference, 1896-1956." In Tsunehiko Yui and Keiichiro Nakagawa eds. Business History of Shipping: Strategy and Structure, University of Tokyo Press, Tokyo. (HE/571/.I67/1985)

Yamey, B.S. (1972), "Predatory Price Cutting: Notes and Comments," *Journal of Law and Economics*:15.

Definition of Variables[45]

Age: years the firm has been established at the time of entry.

Normalized Route Tons: tonnage the firm puts on the new route at the time of entry, adjusted by the length of the route. Normalized Route Tons2=Normalized Route Tons^2.

Firm Tons: tonnage the entrant owns in total. Firm Tons2=Firm Tons^2

Cartel Tons: tonnage operated by members of the cartel in that year.

Big: dummy set to one if Firm Tons > 125,000

Size: *Route Tons/Cartel Tons*, entrant's tonnage size relative to the cartel.

Contract: value in real Pounds Sterling of any contracts the firm has won before entry to carry certain cargo and government subsidies given for specific routes.

Amount Subsidy: Pounds Sterling annual general operating subsidy.

Government Subsidy: dummy variable; 1 if subsidy > 0.

Trade Growth: average of previous two years growth in the sum of real value of imports and exports from the UK to the region.

Trade Opportunities: outside opportunities measured by the growth in value of all UK trade excluding the region. (=%change(real UK trade - real region trade))

Time: a time trend starting in 1875.

Time Since Last Entry: number of years since the last entry in that conference. Undefined for first entrants.

Time Since Last War: number of years since the last price war in that conference. Undefined for first entrants.

Past War Percentage: cumulative percent of previous entries that have resulted in a price war in an observations's conference.

---

[45] The dummy variables below are not orthogonal; an observation's values for *Branch*, *Via the Cape*, and *Big* might all be one, for example.

Table I: Price Movements around Price Wars				
Description	price before	price during	price after	exit?
FEFC forms 1893: textiles to China	30 pence	----	120 pence	---
Indian Mutual war to Calcutta 1891	258 pence	90 pence	NA	Yes
Austrian Lloyd war to Bombay 1881	20 rupees	5 rupees	20 rupees	No
Houston war to South Africa 1902	510 pence	192 pence at the lowest	NA	Yes
NYK war to Yokohama 1896	480 pence	300 pence	384 pence	No

Table II: Expected Signs of Variables under Different Theories of the Causes of Price Wars[46]				
Theory	Long Purse	Renegotiation	Reputation	Cost
Route Tons		+		- (opportunity)
Firm Tons	-			
New	+			
Big	-			
Government Subsidy	-	+		
Contract	-	-		
Publicly Owned	-			
Dispute	-	+		
Trade Growth		-		-
Trade Opportunities		+		+
Past War Percentage			-/?	
Time since Last War			+	-

[46] A blank cell indicates the variable should not have an effect under that theory.

Table IIIa: Descriptive Statistics

Variable	Obs	Mean	Std. Dev	Min	Max
War	52	.3856132	.4865585	0	1
Normalized Route Tons	52	31905.79	31132.49	0	146587
Route Tons	52	34657.34	32920.86	0	151931
Firm Tons	52	100184.5	151628.8	0	836687
Size	52	.0974421	.0902904	0	.5349839
Trade Growth	52	3.500094	13.92096	-45.16	39.42
Trade Opportunities	52	4.006604	16.31162	-27.1	42
Big	52	.245283	.4343722	0	1
New	52	.2264151	.4225158	0	1
Subsidy Dummy	52	.4150943	.4974536	0	1
Contract	52	.3141509	.970645	0	6.65
Publicly Owned	52	.245283	.4343722	0	1
Via the Cape	52	.1509434	.3614196	0	1
Branch Line	52	.7358491	.4450991	0	1
Length of War	51	96.21569	235.9977	0	1230
Past War Percentage	52	.2755	.1721	0	.5385
Time since Last Entry	52	2.188679	2.076022	0	8
Time since Last War	52	3.433962	4.857625	0	25

Table IIIb: Descriptive Statistics for Non-Dispute Entries

Variable	Obs	Mean	Std.Dev.	Min	Max
War	44	.2954545	.4615215	0	1
Normalized Route Tons	44	29349.25	30223.05	0	146587
Route Tons	44	28162.2	25492.43	0	126000
Firm Tons	44	83067.95	118865.9	0	598203
Size	44	.0933082	.0912062	0	.5349839
Trade Growth	44	5.21	13.01451	-12.08	39.42
Trade Opportunities	44	3.871136	16.52959	-27.1	40.6
Big	44	.2272727	.4239151	0	1
New	44	.25	.4380188	0	1
Subsidy Dummy	44	.3863636	.4925448	0	1
Contract	44	.1818182	.3901537	0	1
Publicly Owned	44	.2272727	.4239151	0	1
Via the Cape	44	.1818182	.3901537	0	1
Age	44	30.61364	32.46552	0	135
Branch Line	44	.7045455	.4615215	0	1
Length of War	43	63.86047	180.2323	0	960
Past War Percentage	44	.2718	.1725	0	.5385
Time since Last Entry	44	2.181818	2.049081	0	8
Time since Last War	44	3.636364	5.058223	0	25

Table IV: Determinants of a Price War Conditional Upon Entry Probit Specification			
	All Obs	No Dispute	No Dispute
Constant	-.442 (.558) .433	-.179 (.675) .793	-.802 (.936) .397
Firm Tons	.024 (.014) .100	.037 (.019) .063	.034 (.019) .092
Firm Tons^2	-.049 (.036) .183	-.070 (.052) .190	-.066 (.055) .160
Normalized Route Tons	-.053 (.031) .094	-.135 (.052) .014	-.134 (.053) .016
Normalized Route Tons^2	.516 (.248) .044	1.19 (.428) .009	1.17 (.439) .012
Big	-2.26 (1.19) .063	-4.05 (2.10) .062	-3.91 (2.18) .082
New	1.49 (.616) .020	3.33 (1.10) .004	3.49 (1.17) .005
Trade Growth	-.082 (.039) .043	-.136 (.059) .027	-.142 (.063) .031
Trade Opportunities	.048 (.028) .096	.084 (.043) .060	.093 (.047) .057
Contract	-.071 (.563) .901	-1.75 (.977) .082	-1.50 (.992) .140
Past War Percentage	---	---	2.22 (1.17) .252
Observations	52	44	44
LogLikelihood	-26.24	-14.54	-13.80

(Standard errors are in parentheses; p-values are in the third row.)

Table V: Marginal Effects for the Results in Table IV

	All	No Disputes	No Disputes
Constant	-.125	-.033	-.137
Firm Tons (thousands of tons)	.007	.007	.006
Firm Tons^2 (billions of tons)	-.014	-.013	.011
Normalized Route Tons (thousands of tons)	-.015	-.025	-.023
Normalized Route Tons^2 (billions of tons)	.146	.217	.200
Big (dummy: one if Firm Tons > 125,000)	-.638	-.740	-.669
New (dummy: one if age < 5 years)	.421	.608	.597
Trade Growth (route trade value change)	-.023	-.025	-.024
Trade Opportunities (Rest of U.K. trade value change)	.014	.015	.016
Contract (dummy: one if firm has contract)	-.020	-.320	-.257
Past War Percentage (percent wars out of entries)	---	---	-.380
LogLikelihood	-26.24	-14.54	-13.80

Table VI: Determinants of Entry Tonnage
OLS Regression

	All Obs	No Dispute	No Dispute
Constant	6849 (8498) .425	12943 (7709) .102	14254 (36302) .698
Firm Tons	.073 (.036) .047	.061 (.032) .065	.028 (.039) .478
New	-9850 (9475) .304	-13925 (8567) .113	-26564 (12557) .046
Contract	7764 (9651) .426	14176 (9285) .136	15588 (11573) .192
Trade Growth	1517 (599) .015	1256 (545) .027	969 (647) .148
Trade Opportunities	-767 (495) .129	-511 (456) .270	-399 (527) .456
Past War Percentage	42599 (21176) .051	31422 (19924) .124	56723 (91975) .544
Dispute	28090 (9824) .007	---	---
Time Since Last War	----	---	763 (1036) .469
India Dummy	-4897 (8493) .567	-13204 (8411) .125	-32728 (15815) .050
Far East Dummy	582 (549) .295	582 (495) .247	-15450 (12553) .231
Obs No	52	44	32
Adjusted R^2	.302	.310	.348

(Standard errors are in parentheses; p-values are in the third row.)

Table VII: Comparison of Probit, MLE, and Semiparametric Estimates of the Determinants of Price Wars

	Probit	MLE $\alpha_0 \neq \alpha_1$	Window Width	MRC/IR
α_0, Probability of Misclassification of "No War"	---	.092 (.065)	 0.2 0.5	0 (undefined)
α_1, Probability of Misclassification of "War"	---	3.81E-05 (.696)	 0.2 0.5	0 (undefined)
Constant	-.179 (.675)	-2.23 (1.40)	 0.2 0.5	---
Firm Tons	.037 (.019)	.054 (.036)	 0.2 0.5	.038 (.025) (.070)
Firm Tons^2	-.070 (.052)	-.101 (.093)	 0.2 0.5	-.076 (.051) (.153)
Normalized Route Tons	-.135 (.052)	-.047 (.072)	 0.2 0.5	-.124 (.057) (.169)
Normalized Route Tons^2	1.19 (.428)	-.231 (1.05)	 0.2 0.5	1.07 (.452) (1.34)
Big	-4.05 (2.10)	-6.10 (4.42)	 0.2 0.5	-4.01 (2.34) (6.34)
New	3.33 (1.10)	4.19 (1.95)	 0.2 0.5	2.75 (1.07) (4.48)
Trade Growth	-.136 (.059)	-.170 (.090)	 0.2 0.5	-.122 (.058) (.167)
Trade Opportunities	.084 (.043)	.094 (.060)	 0.2 0.5	.075 (.035) (.122)
Contract	-1.75 (.977)	-1.75 (1.09)		-1.75 (undefined)
Observations	44	44	44	

Notes are on the following page.

Notes for Table VII:

1. Because the estimates of α_0 and α_1 are zero, their standard errors cannot be calculated; an adjacent step is necessary for s.e. estimation.

2. One variable must be held fixed to calculate the standard errors for the Han coefficients. In Table VII it is *Contract*.

3. The MRC results must be scaled in order to be able to compare them to Probit and MLE. The MRC/IR method only estimates the ratios of the coefficients, not their absolute magnitude. I chose a scaling factor which would create equal *Contract* coefficients across the two methods. That scaling factor was then used to scale the other MRC/IR coefficients.

4. The size of the window width used to construct the kernel for computation of the standard errors of the MRC estimates strongly affects the resulting standard errors. As the window width rises, so do the standard errors. The corresponding standard errors in Hausman and Scott Morton (1993) are not sensitive to the window width; they have a sample of 5200 observations. I report results using window widths of 0.1 and 0.5. 0.5 would correspond to the rule of thumb method of Silverman (1986). There is not a general theory that gives an optimal window width in this case, so it is unclear what the correct procedure is.

Appendix

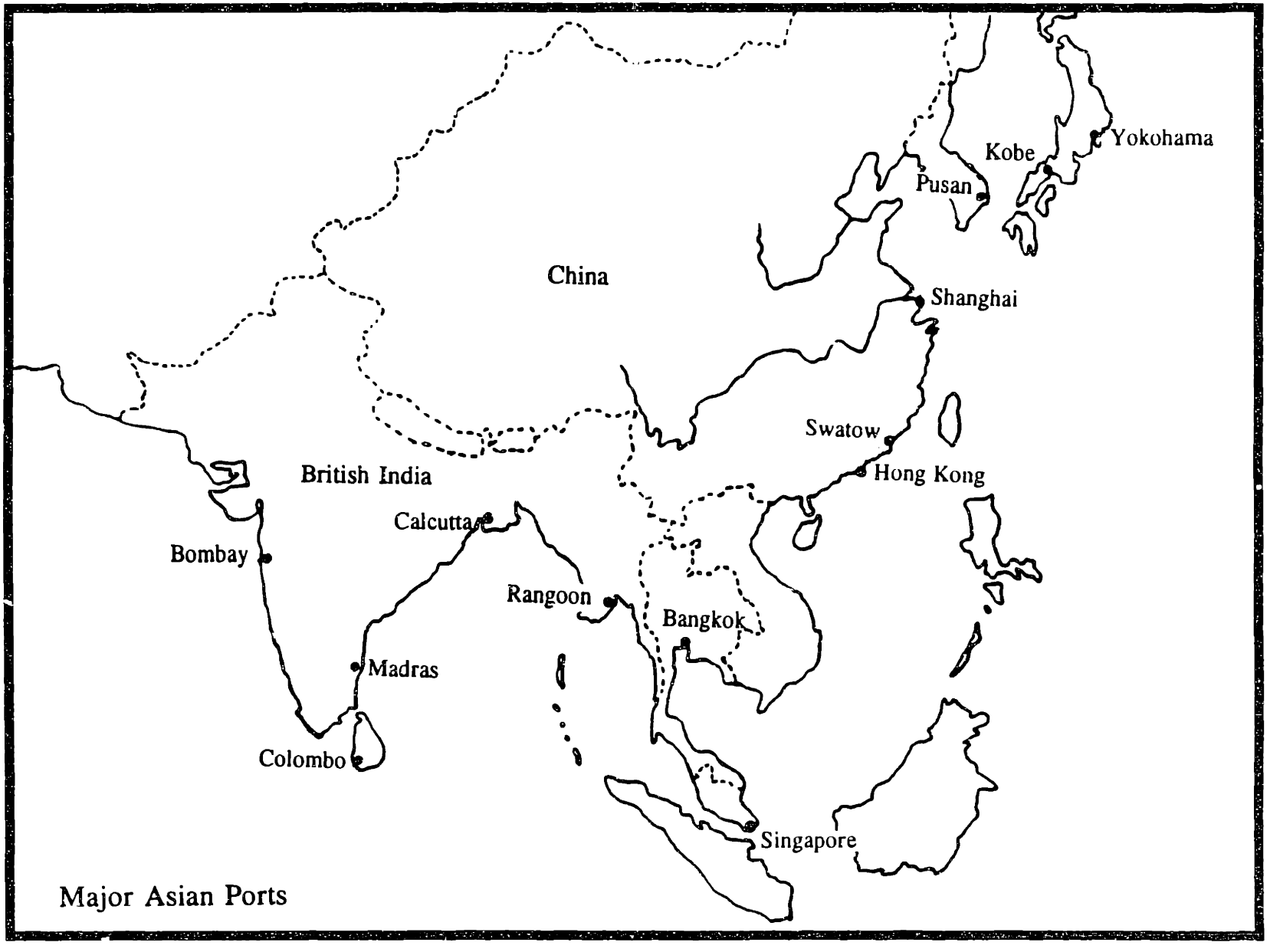
Average Net Voyage Profit in £ '000		
Year	Far East & Australia	Indian & South African
1871	5.1	3.2
1881	4.7	3.9
1885	1.57	.44
1890	.73	1.2
1895	1.9	.76
1900	4.03	1.84
1905	4.29	1.21
1910	3.94	1.74
1913	7.06	3.55

Average Cost per Ton			
year	3,500 tons	5,500 tons	8,500 tons
1885	.0616	----	----
1890	.0570	----	----
1895	.0510	.0501	----
1900	.0486	.0486	.0372
1905	.0440	.0420	.0360
1910	.0450	.0407	.0290
1914	.0412	.0392	.0260

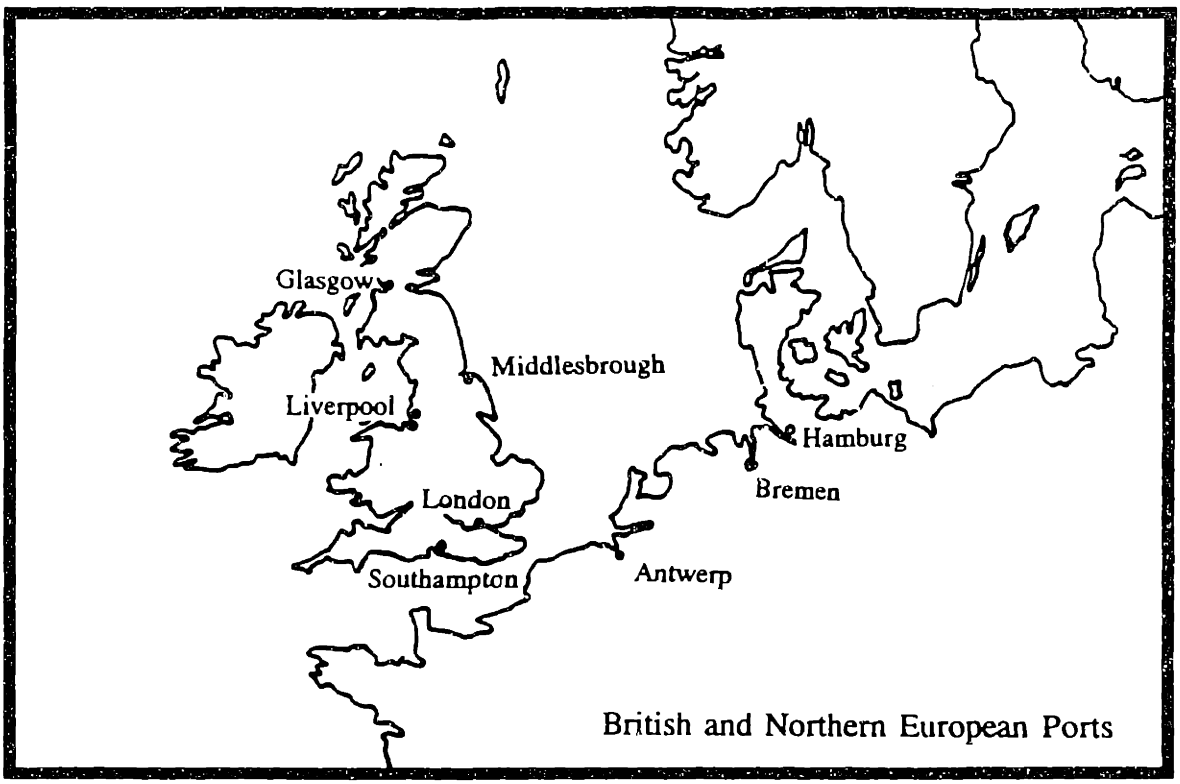
(Information in both tables taken from Hyde, Harrisons...)

Military wars in the regions of interest in the period 1875-1929:
1894-96 Sino-Japanese war.
1899/10 - 1902/5 Boer War in South Africa.
1904-05 Russo-Japanese war
1914-1918 World War I.

Abbreviated names of shipping firms:
DOAL: Deutsche Ost-Afrika Linie
DADG: Deutsche Australische Dampfschiffahrt Gesellschaft
OSS: Ocean Steam Ship Co (colloquially "Blue Funnel")
P&O: Peninsular and Oriental Steam Ship Co
BISN: British India Steam Navigation Co
NDL: Nord Deutscher Lloyd
NYK: Nippon Yusen Kaisha
OSK: Osaka Shosen Kaisha



Major Asian Ports


Glasgow
Middlesbrough
Liverpool
Hamburg
Bremen
London
Southampton
Antwerp
British and Northern European Ports

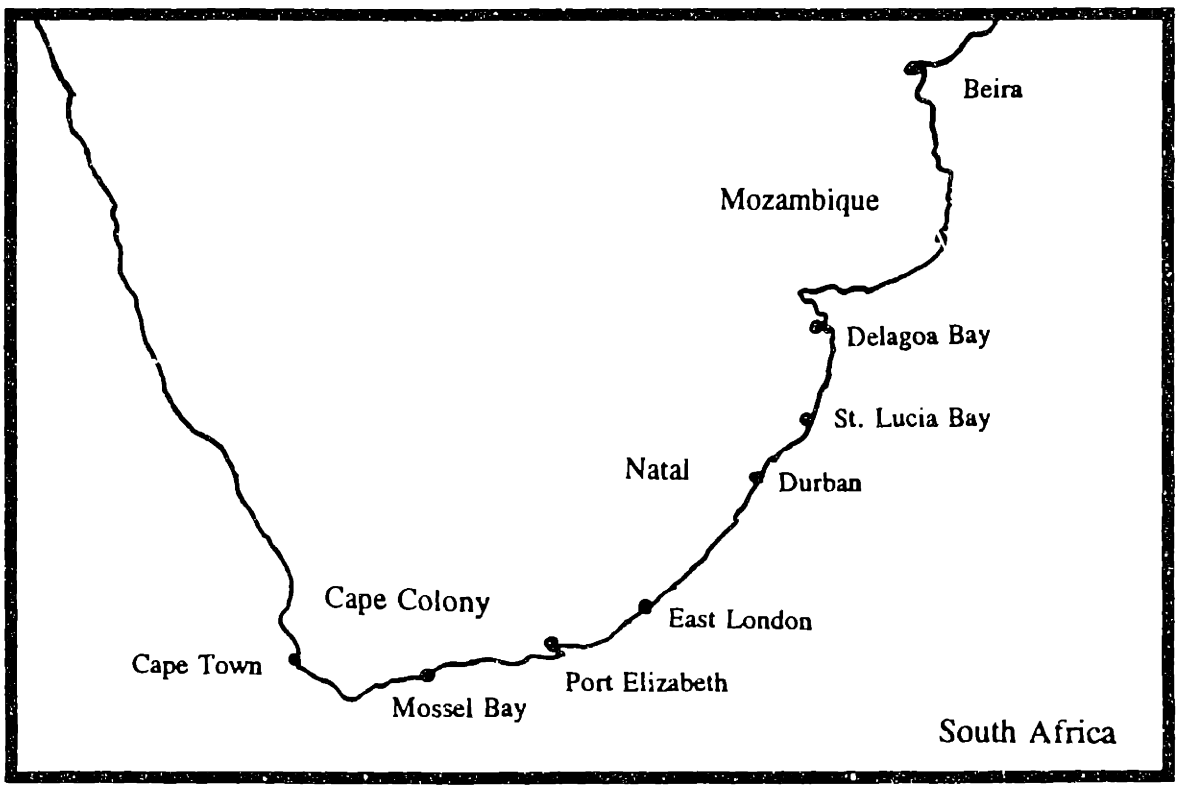

Beira
Mozambique
Delagoa Bay
St. Lucia Bay
Natal
Durban
Cape Colony
East London
Cape Town
Port Elizabeth
Mossel Bay
South Africa

Figure 4

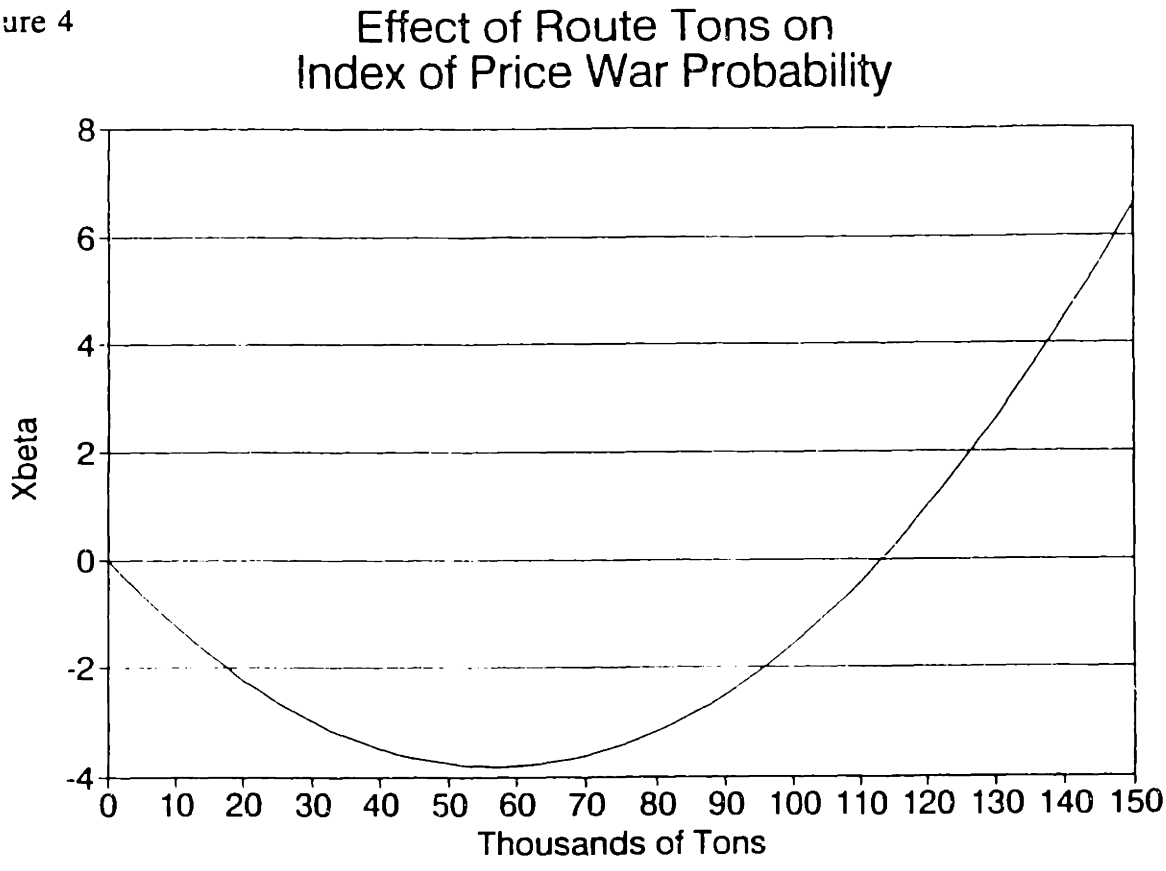
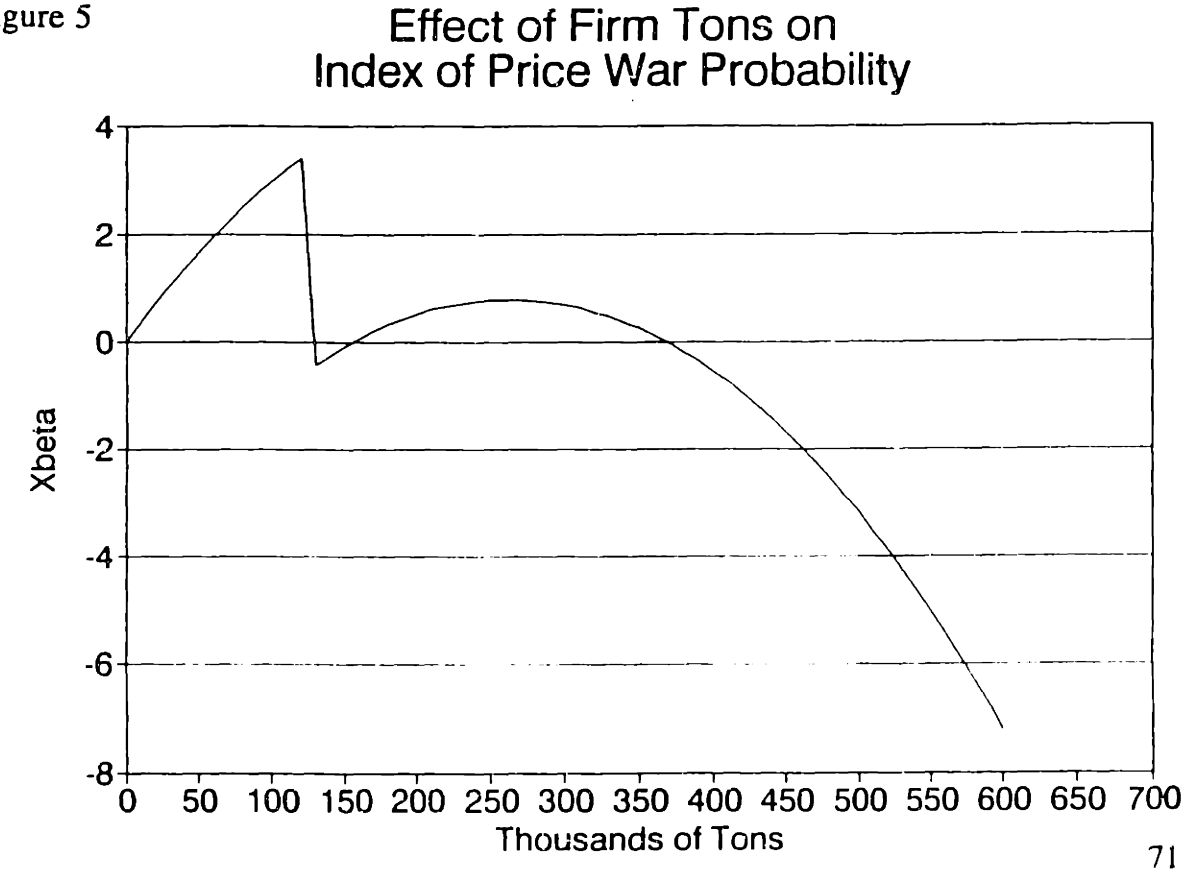


Figure 5



Chapter II

Barriers to Entry, Brand Advertising, and Generic Entry in the U.S. Pharmaceutical Industry

I. Introduction

The question of whether advertising acts as a barrier to entry has been a subject of ongoing controversy in the industrial organization literature.[1] Advertising might disseminate information and help consumers make rational choices. On the other hand, advertising might merely persuade consumers of product differentiation where none exists. The second type of advertising could act as a barrier to entry. Such a barrier to entry will be easier to implement if the incumbent has a long period of legal monopoly with a specific date when entry is permitted, as a patented product does in the pharmaceutical industry. This paper models the entry decisions of generic pharmaceutical manufacturers faced with patent expiration to examine the role of pre-expiration brand advertising.

When a pharmaceutical patent expires, generic firms may enter that market and begin selling an exact replica of the original drug. Market outcomes after patent expiration are heterogeneous; the number of generics entering and the times at which they enter differ for every drug. Additionally, many patented products never experience entry. Such drugs are not expected to generate sufficient revenue into the future to reward entry; perhaps the drug serves a very small market or a closely related product is a superior treatment.

The entry decisions made by many generic firms determine the observed outcome in any given market. Each potential entrant examines the characteristics of the available

[1] e.g. Bain (1956), Dixit (1980), Schmalensee (1982) and (1983).

markets, including those affected by brand behavior such as advertising. The entry decision is based on expected profits, which themselves depend on market characteristics, brand advertising, and the potential entrant's expectations of other generic firms' behavior. For a given revenue stream, a brand producer benefits from fewer generic entrants. The availability of strategies that may be effective in deterring entry is an important issue for producers of branded pharmaceutical products.

Brands have at least two instruments with which to influence the state of the market and generic entry: price and advertising. Limit pricing is a well-known approach that might first occur to an economist thinking about deterrence. A lower branded pharmaceutical price forces a lower generic price; generics cannot sell their product unless they offer a significant price discount. At the time a generic is considering entry, the brand can charge a low price. It may be able to choose a price path and sign contracts that commit it to charge a low price after patent expiration.[2] Generic entrants calculating expected profits over the life of the drug may conclude that the expected profits do not warrant incurring the fixed costs of entry. Alternatively, there could be some uncertainty on the part of potential entrants over the costs of production of a particular drug, or the state of future demand for it, that can be signalled by a limit price.[3]

Agency problems are severe in the pharmaceutical industry. Doctors do not share

---

[2] Hospitals and HMOs routinely receive contracts for one or two years from manufacturers and distributors. Independent pharmacies do not have contracts in general.

[3] Milgrom and Roberts (1982).

in the cost of the prescriptions they write and therefore do not have an incentive to consider the price of different products in making a prescription decision. We might suspect that price is not playing its customary role in this market, and that changing price would have little effect on quantity sold. However, innovator pharmaceutical firms do a great deal of advertising which is primarily directed at the physicians who decide which drugs a patient will use. Therefore the analysis below will illustrate how the manufacturer of a branded pharmaceutical might employ advertising rather than price to prepare for post-patent competition.

This study explores the factors affecting generic firms' decisions to enter markets where patents have recently expired. The pharmaceutical industry is particularly well suited for study because its new products are patent protected. The patent's expiration date is the moment when the market moves from monopoly to potential competition. Additionally, the fact that entrants must satisfy FDA regulations allows me to use a strict and clear definition of a market and of entry.

I model the number of generic firms entering a market as a function of market characteristics and brand advertising. Market characteristics that affect entry include revenue of the brand before patent expiration, elasticity of demand, customer mix, switching costs, FDA regulations, and advertising. I use a sample of 142 drugs that lost patent protection between 1986 and 1992 to examine the relationship between market characteristics, entry amounts and patterns, and advertising. I find that revenue is consistently important in predicting the amount of generic entry, as are dummy variables representing regulatory stringency at the FDA. Increasing the number of consumers with

elastic demand, such as hospital buyers and consumers with chronic diseases, increases generic entry. The factors affecting the number of early entrants are different from those attracting late entrants. One explanation for this result is that there are two types of entrants: those trying to be ready to sell at patent expiration, and those who wait to observe the early outcome before deciding to enter.

Previous work has maintained that advertising plays a role in determining the amount of entry into a market.[4] If advertising previous to patent expiration is exogenous to the entry decisions of generic firms, advertising can be included in the model without instruments. The results under this simple hypothesis indicate that brands that are about to lose patent protection can deter generic entry by additional advertising. However, when I instrument for advertising, it fails to have an entry-deterring effect. My conclusion is that brand advertising is not a barrier to entry by generic firms.

The organization of the paper is as follows. Section II looks at features of demand for prescription pharmaceutical products; Section III discusses the previous literature in this area. The industry and its rules of entry are covered in Section IV as well as a description of my data and data sources. Section V examines trends in price, quantity, and advertising that establish the general pattern of brand behavior. Section VI develops and estimates a reduced form model of generic entry. Concluding remarks follow in Section VII and an appendix presents a simple model.

II. Demand for Prescription Drugs and Agency Problems

[4] Schmalensee (1983) and Spence (1980)

A critical feature of the demand for prescription pharmaceuticals is that the end consumer, the patient, does not select the drug he or she will consume.[5] Instead, the physician picks the drug therapy and also chooses either the brand or generic form. Physicians have no incentive to prescribe the generic version of a drug in the sense of receiving any kind of tangible payoff. Physicians often do not know the generic name for a particular drug and so prescribe the brand, or simply write down the brand name out of habit. A physician might also be risk averse, afraid of lower quality in the generic drug and wanting to avoid the repercussions from a bad outcome. Pharmaceutical firms spend more on advertising, most of which is directed at physicians, than they do on R&D.[6] Advertising emphasizes the brand's therapeutic advances, its absence of side effects, and the quality and service of the brand's firm. Physicians may be uninformed as to prices and availability of generics, since generic firms do not advertise very much.

The majority of prescription drugs, over 60%, are sold through independent drug stores or chain pharmacies.[7] Even if a doctor writes a brand name on the prescription, substituting a generic for the brand may be permitted. Each state in the U.S. regulates generic substitution; it can be forbidden, permitted, or even mandated. (The mandate is overridden if the physician prescribes the brand and indicates substitution is forbidden). The state laws regarding substitution changed considerably in the late 1970s and early 1980s. During this time, most states moved to permit or mandate generic

---

[5] All drugs in my sample require a prescription from a physician before a consumer can purchase them.

[6] Hurwitz and Caves 1988.

[7] National Association of Chain Drugstores.

substitution at the pharmacy level.[8] By intervening between the physician and the patient, the pharmacy exerts an important influence on whether a brand or generic is sold. Usually the pharmacist receives a higher markup rate on generic drugs which compensates him or her for their lower absolute price. Thus pharmacists are in general as disposed towards dispensing generics as toward dispensing branded drugs. The final consumer may be subject to agency problems also; about 40% of prescriptions are covered by some sort of insurance. If the consumer does not pay for the drug, or pays a flat fee, and is at all concerned with quality, he or she may not want to buy a generic even if it is available.

Because of the agency problems already mentioned, price does not have as much influence in the pharmaceutical market as other markets. Instead, advertising is the instrument with which firms can influence demand. However, institutional buyers avoid many of these agency problems. They buy in bulk and have a permanent staff that is up to date and informed about available products. About 30% of prescription pharmaceuticals are dispensed in hospitals and HMOs or by mail order firms. These institutions tend to have formularies, or lists, that restrict the types of drugs available within the hospital or HMO. The formulary committee, composed of physicians, pharmacists, and nurses, has the information and expertise to decide which drugs are needed and are most cost effective. The institution purchases those drugs and others are

[8] In 1993 12 states, including Massachusetts, had mandatory substitution laws. See Judy Hellerstein (forthcoming) for more details. Most of the law changes toward substitution occurred in the late 1970's, before my sample period.

not supplied.[9]

III. Literature Review

The structured competition of the pharmaceutical industry which results from the combination of regulation and patented products has attracted interest from economists. The works below attempt to explain the amount of generic entry while including advertising, but do not treat advertising as an endogenous variable. The overall result is that advertising has an insignificant effect on the amount of generic entry.

Telser (1979) investigated the relationship among entry, price changes, and advertising across therapeutic categories. The data he used are now relatively old, 1963-1972, and include information on how many firms were producing drugs in 17 therapeutic categories over time.[10] He found that entry is positively correlated with advertising and the growth rate of sales, while it is negatively correlated with price changes and the number of firms already in the therapeutic category. However, I expect his results to differ from mine because the regulatory environment has changed so drastically.[11] However, he strongly rejected barriers to entry in the pharmaceutical industry.

Hurwitz and Caves (1988) use advertising, price, and number of generic

---

[9] In exceptional circumstances other drugs may be used.

[10] Generic drugs are not mentioned; they were a very small part of the industry during this time period.

[11] In addition, he examined entry into a therapeutic class rather than into a chemical compound. This requires a model of differentiated, rather than homogeneous, products.

manufacturers to explain market share of a branded drug after it has lost patent protection. They find the market share of the brand increases with its advertising and the number of years it was on patent and decreases with generic advertising. A large price differential between generic and brand decreases brand share as does a larger number of generic suppliers. Their sample consists of 28 multisource drugs observed in both 1978 and 1983. Because the Waxman-Hatch Act[12] had not yet been passed, generic drug approval was relatively difficult compared to today. It is unclear how their sample differs from those of later periods; certainly generic drugs were much rarer and conditions for entry stricter. A supplementary specification seeks to explain the number of generic entrants with sales, advertising, years on patent, and years of competition. They find sales to be a strongly positive and significant cause of entry, but other variables are insignificant. I also find sales to be the major determinant of entry, but I include a richer set of drug characteristics in my estimation which also help to predict entry.

Grabowski and Vernon (1992) use a sample of 18 products achieving over 50 million dollars in sales each at the time of patent expiration. By selecting their sample in this manner, they may find results relevant only to that "size" of drug. They explain the number of generic entrants with price cost markup, number of years on patent, and advertising of the brand. The markup coefficient is significant in their specification, but advertising is not. Their method of calculating marginal cost, taking the asymptote of price as more generic firms enter, could easily lead to bias. Should perfect competition

[12] The Act is explained in more detail in Section IV.

fail in their two year window, (perhaps because the FDA holds up some entrants) fewer firms will enter, price competition will be less fierce, "marginal cost" will be higher, the price cost markup will appear to be smaller, and the model will predict less entry. Endogeneity is clearly a problem in this procedure. I do not try to estimate marginal cost; instead I use revenue as a proxy for profit because marginal costs are usually very small in this industry. The number of firms granted permission to manufacture the drug by the FDA is the correct measure of entry, rather than the number of firms actually putting a label on the product, which is usually different. It is not clear what Grabowsky and Vernon learn by having labelers as the dependent variable; one manufacturer can supply multiple labelers in addition to itself. They also conclude that there is no evidence of entry deterring behavior on the part of incumbents.

Caves, Whinston, and Hurwitz (1991) (hereafter CWH) address a number of these problems in a rigorous fashion. Their data consist of 30 products that lost patent protection between 1976 and 1987. It is worth noting CWH's treatment of two problems that recur in pharmaceutical data: while they weed out patents that did not hold in practice, they include cases where cross-licensing or simultaneous discovery result in two brands for the same patented drug. I follow the same rules when constructing my dataset. Their main concern is to explain price movements, market shares, and quantities sold of both generic and branded drugs. Again, they show that a greater number of generic entrants (instrumented) depresses the generic price and lowers the quantity share of the brand, but they do not model the entry decision directly. The second result of CWH is that brand advertising starts declining two years before patent expiration and

then falls substantially with generic entry. The interpretation they lend to the result is that brand advertising does not limit generic competition post-patent; otherwise it would continue after expiration. In addition, they find that large revenue markets experience a more rapid decrease in advertising following patent expiration than small markets and give a "public good" explanation. However, CWH do not look at whether these patterns differ across different sizes of market and they do not try to explain the generic entry decision other than acknowledging its endogeneity.

IV. The Pharmaceutical Market

Incentives to innovate and discover new "ethical" (prescription-only) pharmaceuticals depend importantly on patent rights and the associated profits captured by the innovator. Patents are granted for several types of discovery, including new chemical entities, new processes (manufacturing), and new devices (delivery systems like IV). The patent holder can exclude anyone from producing an identical product; the process patent holder can exclude anyone from employing the patented process. Exclusivity rights to the invention last for 17 years, whereupon it becomes public property. During its exclusivity period the monopolist earns rents and accumulates goodwill or reputation, and prepares for a known date when it will face competition.[13]

[13] Finding the effective patent expiration date can be complicated. For example, a process patent may be binding after the patent on the chemical entity has expired. Patents on pharmaceutical processes and devices are not monitored by the FDA; the determination that a competitor has violated the binding patent is made in court. Additionally, any issue involving an antibiotic patent is resolved in the courts by suits among an innovator and generic manufacturers; the FDA does not monitor patents in that class of drugs.

Manufacturers fall into two main categories. The first category contains "pioneer" firms; they undertake research and development to discover new drugs and bring them to market. The total cost of developing and testing a new drug in the U.S. is now over $120 million.[14] In order legally to produce and sell the new product in the United States, a pioneer firm must first receive a New Drug Application (NDA). An NDA requires expensive clinical testing, which may take a number of years, in order to prove safety and efficacy to the satisfaction of the FDA. The years of testing reduce a drug's ability to earn monopoly profits by eroding time under patent protection. The Patent Term Restoration Act (Waxman-Hatch) of 1984 was designed to restore some patent protected years lost to FDA testing. A drug's patent can be extended for up to five years, less for minor patents.[15]

The second type of firm is a generic or imitator firm. After patent expiration, any firm may apply for an Abbreviated New Drug Application, ANDA. Since 1984, generic manufacturers are no longer required to duplicate the original safety and efficacy tests performed by the brand, which had raised entry costs. Instead, the generic firm must show its product to be "bioequivalent" to the original branded product. The standard test involves giving the generic drug to healthy human subjects and testing quantities and rates of absorption into the bloodstream and at the therapeutic site. The exact application requirements differ by drug; in general more expensive or difficult tests

---

[14] F.M.Scherer (1993)

[15] The FDA grants exclusivity periods in which it will not approve another firm to manufacture a drug, for any one of a list of innovations. Firms that have patents for a new strength, new form, new route, or new combination can be eligible for patent extension.

are required for drugs that are newer and less well understood or are not absorbed by the body in well-known ways. In addition, FDA inspectors examine the equipment the firm plans to use to manufacture the drug, and inspect early batches of the drug. The cost of submitting an ANDA in the early 1990s was about one million dollars, dramatically lower than the cost of an NDA.

Should a generic firm submit an ANDA to the FDA, that fact is confidential information. The firm may release the information, but usually does not; the FDA will not release the information at all until such time as the ANDA might be approved. Thus, many firms are effectively moving "simultaneously" because they do not see each others' actions. The ANDA application process takes about eighteen months from first submission of the application to final granting of permission.[16] The question of whether generic firms supply as high quality a product as the branded firm is a matter of debate. The FDA tries to keep the generic standard high with surprise inspections and random testing. However, events such as the generic scandal of 1989 suggest that some generic firms could be producing a lower quality product. In that year it was discovered that some generic firms had submitted samples of the branded drug to laboratories for bioequivalence testing instead of their own product. In addition, some generic firms were bribing officials at the FDA to speed approval of their applications. Variation in time to approval occurs because firms' applications differ in quality and because the FDA is an unpredictable bureaucracy. Because of long processing time for applications, other firms may have decided whether or not to enter before a first mover's ANDA is granted.

[16] The length of time to approval has been growing over time.

Data

A "drug" refers to a specific chemical entity that may be called by its generic or brand (proprietary) name. A drug may be manufactured by either the NDA holder or one of many ANDA holders. A firm decides what forms and concentrations to produce; a different (A)NDA is required for each. In my data I characterize forms as solid (e.g. tablets or capsules), injectable, or other (syrup, cream, patch, inhaler). Very different manufacturing equipment is needed for the different forms. Within a form, a drug can come in different concentrations (e.g. 250 mg or 500mg tablets). A unit of observation is a drug-form-concentration combination.[17] There are 113 drugs, 142 drug-form combinations, and 252 drug-form-concentration observations in my dataset. Table I notes the number of observations that fall into each form category. The drugs that experienced patent expiration in my period belong to a variety of therapeutic classes as tabulated in Table II.

Data from IMS America

Revenue and quantity data were supplied by IMS America from their Drugstore and Hospital Audits.[18] I observe revenue and quantity in April, August, and December from two years before patent expiration to one year after expiration for all the drugs that lost patent protectior between 1986 and 1991. I also collected monthly advertising data

---

[17] A more complete description of each observation is included in my original dataset: chemical, labeler, form, concentration, and quantity are listed.

[18] IMS America is a firm which collects and processes different types of pharmaceutical data and sells the information to customers, most of whom are pharmaceutical firms.

from three years before, to one year after, patent expiration. IMS records the monthly expenditure on journal advertising and detailing for each drug-manufacturer pair. Each advertising expenditure observation therefore encompasses all concentrations of a drug.[19] Detailing is the process of drug company representatives making office visits to doctors and presenting information about certain drugs in person. Printed advertising is published in medical journals that physicians read. Spending on journal advertising is about a quarter of the amount spent on detailing.[20]

The dollar figures were deflated by the Urban CPI to make prices comparable across years. I rearranged the data into nine four-month periods where periods 1-6 contain the data for the two years before patent expiration and periods 7-9 have data for the year after patent expiration. The exact date of patent expiration falls in period 7 for each drug, regardless of calendar date of the expiration.

Data from the FDA

I compiled information from the FDA's *Approved Drug Products* giving the number of ANDAs granted through zero, twelve, or twenty-four months after patent expiration. These variables are named *Early, Middle*, and *Total*. The mean number of firms entering a market within one year after patent expiration (*Middle*) is 0.99. A substantial number of markets, 104 out of 142, experienced no entry within one year

[19] Occasionally IMS divides up a product by form or concentration. For example, the extended release version of a drug is usually a different concentration and might have advertising expenditures that appear separately.

[20] Journal expenditures are measured more accurately because IMS subscribes to every existing medical (including nursing and dental) publication and counts and estimates the cost of the advertisements therein.

following patent expiration.

One problem with measuring approved ANDAs is that a firm may have expected to receive its ANDA at date t and encountered unexpected FDA delays. The months following the generic scandal were just such an occasion; many staff members of the FDA's Generic Drug Division were fired and others became very cautious and slow in approving applications. To control for this problem, I obtained from the FDA the application dates for all approved ANDAs. Instead of looking at which firms were granted ANDAs near patent expiration, I measure how many firms applied for that ANDA at some point in time. Unfortunately, application dates for applications that were not ultimately approved are not included in the publicly available dataset. I constructed variables measuring how many firms applied (and were eventually approved) for a specific ANDA by 18 months preceding patent expiration, by six months before expiration, and not later than six months after expiration. These variables are named *Applic-Early*, *Applic-Middle*, and *Applic-Total* respectively.[21] I also define two variables that measure latecomers to the market both in applications and approvals. *Applic-Late* is defined to be (*Applic-Total*) - (*Applic-Middle*). It is the number of applications submitted between six months before and six months after patent expiration. *Late*, being *Total* - *Middle*, should capture a mixture of bureaucratic delay and purposefully late entrants. Notice that all *Early* entries are contained in *Middle* which is itself contained in *Total*. *Early* or *Middle* and *Late* are non-intersecting sets.

---

[21] A firm with an ANDA approval is not required to produce the drug; the firm will not lose its ANDA due to non-production. It is free to manufacture the drug again after a period of non-production.

The average length of the approval process over all years and products in my data is 18 months. The generic scandal significantly raised approval times; all years since 1989 have higher average approval lengths than those before the scandal (this also may be affected by the composition of drugs going through the process in any given year). Average time to approval varies widely by drug and firm. Firm specific approval times calculated from this dataset vary widely. Some generic firms may have better quality management or may perhaps choose drugs that are straightforward to test.[22] In this paper I analyze both approval and application dates; since approval time is affected by the amount of effort and resources put into the application, the number of early approvals, for example, reveals some information about the desirability of entering a given market which is not contained in the early application measure. In addition, if a drug is known to have a short approval time, interested firms will submit ANDAs relatively late; in that case early approvals will be a better measure of the amount of generic entry than early applications.

Table V shows the distribution of entry over time. The measures that are 18 months apart show considerable similarity: see *Applic-Middle* and *Middle*, for example. I have complete information on approvals of ANDAs through July, 1993, so all measures that can be completed through that date have been included.[23] Patents that expire in 1992 or late 1991 have missing values for total and middle applications and approvals

---

[22] Another explanation could be that these firms successfully bribed FDA officials. I have not yet undertaken a rigorous analysis of the data to explore these possibilities.

[23] It is possible a very slow firm applied 18 months before a 1992 patent expiration and had not been granted approval by July, 1993. However, I assume this is not the case.

because not enough time has passed to allow construction of the variables. Since all the application dates I have represent ANDAs that were eventually approved, the entry and application variables are correlated; the correlations range from .42 to .82. Matching all datasets eliminated drugs that were not defined consistently across data sources or were omitted from one or more sources. There remains 142 usable patents and exclusivity rights.[24]

The Decision to Enter

A generic firm deciding whether to enter a pharmaceutical market has to consider issues of demand, supply, timing, and expectations. At any given time there is a stock of expired patents and soon-to-be expired patents from which a generic firm could choose. More-recently- or not-yet-expired patents will in general describe more advanced and efficacious drugs which will be more popular therapies. The number of patents expiring in a year can vary considerably as indicated by the summary statistics in Table IV.

The generic firm must make a judgement about the future profitability of the compounds available. For example, the demand for a specific drug might grow over time as the population ages. Therapeutic substitutes for a specific drug might not exist, or they might be about to lose patent protection themselves. A pre-existing drug could be found to have new properties, increasing the demand for it.[25] The price elasticities

---

[24] It is possible that I have included patents that have expired but for which a process patent or some other unknown feature is preventing entry by generic firms. I have no way to distinguish these cases, if they exist.

[25] For example, ulcer drugs are also good antacids.

of the drug's consumers will determine how readily they will buy a generic. Switching costs (if a patient is already using the brand) are one of the factors determining how quickly the generic gains market share. Another important feature is the technological difficulty of manufacturing the drug and receiving FDA approval.

Future supply and markup are also critical; if many generic firms are approved to make the same drug, margins will be driven down in subsequent price competition, and that market may not have been worth entering *ex post*. Firms can decide to abandon production of an "over-entered" drug; in that case the sunk cost of entry has been wasted.[26] A firm's expected markup, and therefore interest in entering a given market, is inversely related to the number of other applicants.[27] When making its entry decision the firm uses all the information it has about the decisions of other entrants. Because the ANDA approval process takes on average 18 months, a firm can only know of applications submitted over a year earlier and regulatory delay makes even that knowledge imperfect. However, entrants submitting ANDAs between six months before expiration and six months after probably have knowledge of approvals also made in that window.[28]

---

[26] Low marginal costs of production tend to make the firm want to continue to produce even as the price falls.

[27] Other applicants might provide information about the state of future demand for the product and thus generate some positive correlation with the firm's desire to enter.

[28] Rumors of pending ANDAs may circulate, especially as expiration approaches, so any estimate of the information structure cannot be too precise. According to sources in the industry, submitted ANDAs remain secret until close to patent expiration, when the firm starts to advertise its future product. The major source of leaks are suppliers of bulk chemicals. An ANDA must list the firm's input suppliers, so a supplier could be aware of the existence of other entrants.

If a firm submits its ANDA in time to be approved by patent expiration, it must make its entry decision with two years still to run on the patent. At that time it is unlikely to know what other firms are also entering. In this sense the firms' decisions are simultaneous and independent. The simultaneous entry game played at patent expiration may not have resulted in the optimal number of entrants. An entrant which enters late enough to observe that outcome may have the chance to profit from filling in gaps. This line of reasoning suggests there may be two types of strategies for generics. The first strategy is to enter early. The firm applies at a time such that it expects to receive approval before the patent expires.[29] The first generic firm to enter a market can set a price that is significantly above marginal cost and yet significantly below the brand's price. Profits earned by the first entrant in its period of "duopoly" drive the race to enter first. Late entrants can actually see the outcome of the early game before submitting their ANDA and still expect to have an approval 24 months after patent expiration.[30] These later firms can react to "mistakes" in the equilibrium outcome.[31]

The generic firm must also decide what strengths to make as it is choosing which compound to enter. The ANDA requires proof that the firm owns the appropriate equipment and has contracts with reputable suppliers for the ingredients. Therefore the marginal cost of applying for an additional strength is low. However, a separate

---

[29] Very early applications (4 years before expiration for example) are generally not approved by the FDA until fairly close to expiration.

[30] Later applications may get approved faster as the FDA gains experience with the drug. I plan to analyze approval times more thoroughly in later work.

[31] However, such a late entrant might be subject to limit pricing or other strategic behavior by the first generic entrant.

application must be submitted and appropriate lab work done and batches tested for each concentration, so concentration entry has a positive cost. In my dataset generic firms apply for all the concentrations of a drug in most instances.[32] To avoid weighting the data by the number of concentrations associated with each drug, I do not include a separate observation for each concentration. Instead, only the concentration with the highest revenue is included in the dataset used for estimation.

V. Price, Quantity, and Advertising Trends

In this section I examine the shape of brand journal advertising expenditures around the patent expiration date. For example, a brand might engage in a boost of advertising at some point close to expiration, let expenditures gradually decline, or suddenly stop promotions at the expiration date.[33] I transform my sample of 142 highest revenue drug-form-concentrations observations into 142*48 monthly observations of journal advertising expenditure. Table VII shows the simplest specification explaining log advertising with a time trend, a shift effective only after patent expiration, and a time trend effective after patent expiration. Each drug has its own intercept.[34] Columns one

[32] Often the firm is in the process of applying for all concentrations but does not time its applications simultaneously. In other instances the avoided concentration might be one of five or six or have very low sales.

[33] Although I have data on detailing, I do not use them here because the original data are recorded in minutes rather than dollars and are subject to considerable measurement error.

[34] I add one to each observation of advertising expenditure in order to be able to work with the log of advertising; several drugs only have expenditures equal to zero. Although it introduces a small amount of bias, I can examine percentage changes over time without losing a substantial fraction of my observations.

through three of Table VII show that overall monthly growth of real spending on journal advertising is perhaps slightly negative, but not much different from zero. However, after patent expiration causes a drop in advertising expenditure of about 15-19%, while the time trend after expiration is approximately -2% per month. Column three reports a pre-expiration downward shift of 14% in advertising expenditures, consistent with Caves, Whinston and Hurwitz's finding.

It may be the case that advertising paths differ by "size" of drug; large revenue drugs may have different trends in advertising than small revenue drugs. To examine the differences among groups, I divide the sample into four groups by revenue before patent expiration: very small, small, medium, and large. The divisions are detailed in Table VIII. Each group gets a separate shift dummy for the year after patent expiration and the year before as well as its own time trend. In addition, every drug has its own intercept in the regression reported in Table IX. Larger revenue drugs are decreasing advertising less in percentage terms than smaller revenue drugs over the whole period. However, large revenue drugs drop advertising substantially after patent expiration, whether measured in a shift or a slope. The year before expiration, very small drugs increase advertising and very large drugs decrease advertising;[35] the middle two groups do not change advertising significantly. The large revenue drugs may be expecting lots of entry and may therefore want to reduce expenditure on advertising that will soon be

[35] Multiple expiration dates within a brand name cause aggregate advertising patterns to become hard to explain. For example, one strength of a brand might be being introduced while another is maturing, and only the aggregate advertising figure will be reported. Restricting the sample to brands with no products already off patent doubles the coefficient on the pre-expiration jump for the smallest group and raises the t-statistic to 2.5.

a public good. The final specification adds characteristics of the market in post-expiration dummy form to help explain advertising movements. Duopoly markets display an additional decrease in advertising of 22% after expiration, and each additional entrant at expiration is associated with a 15% fall in post-expiration advertising.

Brand price and quantity trends are analyzed with a time series dataset also constructed from the cross-section. Nine periods are available for each cross-section observation; price and quantity are observed for drugstores and hospitals separately. I include dummy variables for the year after patent expiration as well as a time trend for each group. The groups are again divided by revenue into four categories. The results are presented in Table X. Brand prices do not react to patent expiration at all for almost all groups. This seemingly odd statistical regularity confirms Caves, Whinston, and Hurwitz results; brand prices do not decline much if at all in response to generic entry. Drugstore prices are trending upwards by zero to three percent every four months, while hospital prices are not increasing at all over time. Quantity declines markedly after patent expiration for the highest revenue drugs. The result is not surprising as these drugs experience most generic entry.[36]

VI. An Empirical Model of Entry

The most important factor in predicting entry is the total profit of the product

[36] I analyzed the advertising/revenue ratio in this specification, but neither trend nor shift coefficients were significant. Part of the problem is that neither measure of advertising or revenue includes 100% of the data. The trend of either measure is not misleading, but the ratio may be quite badly mismeasured and therefore uninformative.

before patent expiration; profit is highly correlated with revenue when marginal costs are low, as they are in the pharmaceutical industry. I created the variable *Ln Revenue* by taking the log of the sum of the dollar revenue, measured in tens of millions of dollars, of all strengths and quantities within one chemical-form (in other words, summed over concentration) during the 12 months before patent expiration. An additional factor that will affect entry is the form of the product. Generic firms usually specialize in producing injectables or non-injectables. Injectables require special equipment from solids and the firm must pass FDA sterilization tests and inspections. Also, the capital equipment required to produce solids is different from that required for other forms. Technical characteristics of a particular form alters the number of potential entrants, and perhaps the number of observed entrants. Two dummy variables, *Injectable* and *Other Forms*, control for form relative to the omitted category, solids. The dummy *Other Form* refers to a few products like patches, tape coated with a drug, liquids applied topically, and inhalers, types of products that do not belong in either the solid or injectable categories.

Share Hospital is the revenue share of that drug-form sold to hospitals rather than drugstores. Unfortunately, the effect of share sold to hospitals and the effect of the different form, injectable, cannot be separated in this dataset. The mean (weighted) hospital share for injectables is 90% (.21) while it is 7% (.12) for non-injectables. If hospitals are more willing to buy generics, a market with a large *Share Hospital* will be more attractive to entry, all else equal. Due to colinearity I omit the injectable dummy from all regressions hereafter.

Duopoly is a dummy variable given the value one when two firms make a branded product under the same patent. Duopoly markets are an unusual feature of the pharmaceutical industry. For one of a variety of reasons, two firms both have rights to the same patent. The products they produce are chemically identical, but have different brand names. Observations in this category (13 out of 142 observations) sum revenue and advertising values for both of the individual firms. If the duopoly has advertised the drug twice as much, the market may be "bigger" in some sense not captured by revenue measures. On the other hand, a duopoly means an additional firm is competing in the post-patent market which would suggest the market is less attractive to enter. Duopolies are not typical markets; the fact a duopoly exists may be due to some feature which also affects generic entry.

An important characteristic affecting entry is the extent of therapeutic substitution available for a particular drug. If there exists an off-patent therapeutic substitute that is a better treatment than the drug in question, fewer generic firms will be interested in entering the market. The timing of patent expirations of a drug's therapeutic substitutes will also affect how many generic firms enter. Therapeutic substitution is a difficult attribute to measure; therapeutically "close" drugs are good substitutes on some occasions for some patients, and not for others. It is therefore difficult to find a good measure of therapeutic substitution. A reasonable variable might be the number of brands in IMS' definition of therapeutic class.[37] I count the number of brands in a drug's therapeutic class and record their patent status. The number of already expired brands should affect

[37] IMS' categories are somewhat arbitrary, but largely consistent.

the amount of generic entry negatively, but a linear form is not the best choice. For example, the effect on entry caused by a rise in the number of off-patent brands from 9 to 10 may not be the same as from 1 to 2. I define *Substitutes Off Patent* to be a dummy variable taking the value one when the number of already expired brands is greater than two. The number of substitutes on patent, which measures the remaining opportunities in the class, never significantly affects entry and is not included in the model.

I also determined which drugs in my dataset treated *Chronic* or acute conditions.[38] To the extent the patient has input about the choice of drug, medications for chronic diseases should have more price elastic consumers; repeat purchasers have greater opportunity to collect information about prices and substitutes. On the other hand, patients being treated for a chronic condition with a branded drug may not want to switch drugs. Their physician may be satisfied with the therapy and reluctant to risk adjustment costs incurred by switching to a generic. Patients who have prescription drug insurance or pay a flat fee per prescription have no incentive to leave the brand. Advertising expenditure is measured very simply by adding up expenditures on journal advertising during the three years before patent expiration. *Advertising* is the sum of journal advertising and detailing expenditures in the third and second year before patent expiration. *Advertising* is normalized to lie between zero and one which requires its units to be 50 millions of dollars.[39] Finally, I include year fixed effects; a drug is identified

[38] Medical reference texts provided most information. David Bailey of MIT's Health Services was very helpful.

[39] Convergence is easier with normalized advertising.

by year of patent expiration. I am particularly concerned with capturing the effects of the generic scandal of 1989 and the subsequent decline in ANDA approvals because of an overhaul of the Generic Drug Division of the FDA. Table IV presents descriptive statistics for all independent and dependent variables.

I first maintain the irrelevance of brand advertising for generic entry decisions. Previous studies of this phenomena (Telser (1979), Hurwitz and Caves (1988), Grabowski and Vernon (1992)) found no influence of advertising on the number of generic entrants. I employ a Poisson count model to explain the entry decisions of generic manufacturers, but do not include any advertising measure as an explanatory variable. This specification will yield consistent estimates if the hypothesis maintained in the previous literature is correct and it provides a convenient benchmark. If advertising is an omitted variable in the determinants of entry, the coefficients in this regression will be biased. Table XII and Table XIII report the results of these specifications. I have a some overdispersion in my data; the variance of an entry measure is larger than its mean. However, I have corrected the standard errors in the reported regressions to allow for misspecification of the conditional variance as in Wooldridge (1991).

I use a standard Poisson specification where $E(N_i|X_i) = \lambda_i$. λ_i is parametrized as follows:

$$\lambda_i = \exp(\beta_0 + \beta_1 LnRevenue_i + \beta_2 Duopoly * LnRevenue_i + \beta_3 Duopoly_i + \beta_4 Chronic_i + \beta_5 ShareHospital_i + \beta_6 OtherForms_i + \beta_7 SubstitutesOffPatent_i + \beta_8 NumberConcentrations + \sum_{t=1987}^{1991} \alpha_t Year_{it}) \quad (2)$$

The coefficients of a Poisson regression do not provide estimates of the marginal effect of changing an explanatory variable. When using a Poisson functional form, the marginal effect of variable i is given by β_i mean(exp(Xβ)). Marginal effects for all variables are provided in Table XIV. For example, the effect on entry of a dummy variable changing value from zero to one is simply its marginal effect.

There is surprisingly little early entry, or approval before patent expiration, in the time period covered by this dataset. Unexpected FDA-caused delays, such as learning about the new Act in the first part of the sample and the generic scandal in the second part, may be part of the explanation for lack of early entry; in that case *Middle* and *Applic-Middle* will represent the group of firms attempting to enter at patent expiration, but exclude those entrants which are purposefully late. *Total* and *Applic-Total* measure the full amount of entry within a two year window without regard to timing. These four dependent variables are most important for the model and form the core of the results. I first discuss the results for these dependent variables which are in columns 2 and 3 of Table XII and Table XIII.

Ln Revenue is significant in explaining entry across all four dependent variables. The predicted number of entrants rises by about 1-1.5 if a drug with ten million dollars

of revenue doubles its revenue and its other characteristics stay fixed. The coefficient on *Duopoly*Ln Revenue* is significant and negative. A duopoly market attracts a smaller amount of generic entry as its revenue increases when compared to an otherwise identical market. One explanation for this result is that competition after expiration must be stronger because of the additional firm in the market. *Share Hospital*'s coefficient is positive and significant in three out of four cases. Recall that it is capturing the effect of both hospital share and the factors associated with injectable production or demand. If hospital share alone were determining the coefficient, the results would imply that an otherwise average drug with 100% hospital share attracts over one more entrant than an equivalent drug with 0% hospital share. The existence of several off-patent substitutes (*Substitutes Off Patent*) in the brand's therapeutic class does not appear to affect the amount of generic entry.[40] I conclude from this that it is very difficult to construct a meaningful measure of therapeutic substitutes. However, the coefficient on *Substitutes Off Patent* is always negative, which is at least consistent with the theory, and is significant for late applicants. Dummy variables for zero or one substitute did not perform any better; neither did *Substitutes On Patent*. *Other Forms* consistently attract less entry due to higher FDA approval costs. One to two firms do not enter the average market, conditional on revenue, because a drug is not an oral solid. *Chronic* has a positive and significant coefficient in all but one specification; the greater elasticity of demand of "chronic" consumers makes chronic markets more attractive, but the

[40] The variable *Substitutes Off Patent* was the only one to lose significance after robust standard errors were calculated.

unwillingness of customers to switch drugs works in the opposite direction. It appears from these results that the elasticity effect of *Chronic* is stronger.

Purposefully late entrants submit an application after knowing more about the market outcome. The motivations of this group may differ from middle and early entrants, so I want to analyze it separately. *Late* includes firms experiencing regulatory delay as well as those that are purposefully late so it is not as clean a measure as *Applic-Late*. *Number of Concentrations* is insignificant in predicting the total amount of entry, but is significantly positive in explaining late applications. More concentrations mean higher costs, conditional on revenue, and therefore entry into a drug with lots of concentrations is less desirable. Perhaps late entrants must fill less desirable markets or perhaps more concentrations slows down the firm's application process. *Chronic* does not affect the amount of successful early entry. Profits from being the only generic for a few months drive the race to be first in a market; chronic consumers may not switch to the generic quickly enough to generate those profits. *Late applicants* do not seem to be affected by duopoly markets.

Notice that the year dummies are significant and have the expected sign for each cohort of applicants. The results of the generic scandal are evident in negative coefficients on *Dum90* and *Dum91* for *Early* firms and *Dum87* and *Dum88* for *Late* firms. Both groups would have been slowed down by the FDA in early 1990. In the application specifications, where we expect the generic scandal to be less important in determining the number of "entrants," the year dummies are much less large and significant.

The Addition of Advertising as an Explanatory Variable

How should we expect advertising to affect the amount of generic entry? The appendix contains one possible model which illustrates how consumer preferences affect advertising, pricing, and entry. Close to patent expiration or afterwards, the goal of advertising is to convince consumers to purchase the brand rather than the generic. If enough demand is shifted to the brand, little will remain for the generic to serve. Sufficiently small expected generic demand will discourage generic entry. The result of the model is that brand firms can choose a profit-maximizing, pre-expiration advertising level which then determines optimal brand and generic prices and the amount of generic entry. I construct a discontinuity by making the hospital/HMO market unaffected by brand advertising. Additionally, I impose the simplifying assumption of a linear demand curve and Cournot competition on the part of generics. In the resulting framework, the brand can deter entry if the market is small enough or if hospital/HMO expenditures are below a critical level. The brand may not choose to deter entry, even if it is physically possible to do so, if the costs are sufficiently high. Deterrence is not worthwhile when it requires so much advertising expenditure that profits end up being lower than if entry were permitted.

A traditional interpretation of advertising is that it expands the market for the product. The market-expansion role of pharmaceutical advertising applies principally at the beginning of the product life-cycle when the drug is new to doctors. The detailing and journal advertising figures I have show expenditures an order of magnitude or more

higher at introduction of a new product than they are later. By the end of the patent life, advertising is only expanding the market to the extent it is defending a drug's share against its therapeutic substitutes. A greater share of a therapeutic market brought about by advertising might encourage generic entry. The question to be answered is therefore, how does advertising affect entry? Does it deter or encourage entry and does that effect differ by characteristics of the drug and market?

Advertising is determined jointly with entry and price in the model above. If the conditional mean of the number of entrants is specified correctly, advertising may be exogenous. That is, if no variables that predict entry and also predict advertising are omitted, then including advertising as an explanatory variable will result in consistent estimates of the coefficients. If not, then including advertising will result in biased coefficients. Regressions including advertising are presented in Table XV; for simplicity, only total and middle dependent variables are analyzed. Advertising has a negative and significant effect in three out of four cases. The advertising variable is normalized to lie between zero and one although the highest advertising level in the dataset is about 50 million dollars. A coefficient of -1.5 is small; advertising may be deterring entry, but the effect on the expected number of entrants of increasing advertising is minimal. The results of this regression are qualitatively similar to those of previous studies that concluded advertising is not a barrier to entry.

Making the assumption that there are no omitted variables in the entry specification that determine the level of advertising is overly optimistic; it is just such drug-specific variation that the econometrician will not be able to include in the

specification. Therefore, we might have strong priors that advertising is endogenous. In the next set of regression results advertising is instrumented in order to estimate a consistent effect of advertising on the amount of generic entry. The instruments I use include the amount of time the drug has spent on patent, which is a measure of where expiration falls in the lifecycle of the drug. Another instrument is a dummy variable for those instances when an advertising observation includes expenditure on a product that has the same brand name as a product in my dataset but a later expiration date. Advertising expenditure will be different for those observations since expenditure is really an aggregate of expenditure over two or more products. The manufacturer's total detailing expenditure for the year is the third instrument. Firms that have big detailing forces in operation for promoting other drugs will find it less costly to undertake some marginal advertising for the drug losing patent protection. The final instrument is the number of physicians that could be expected to use a particular drug. This measure is simply the total number of physicians that report themselves in a one of a number of subspecialties that would use the drug. It indicates how costly advertising the drug will be; journal and detailing expenditures must be higher to reach more physicians. A regression of *Advertising* on only these four instruments and a constant results in an R-squared of 0.13.

The most appropriate technique for estimating this regression with an endogenous variable is Nonlinear Two Stage Least Squares. The nonlinear function is simply the exponential, so the equations remain in a Poisson framework. The results are reported in Table XVI. All coefficients decrease in significance compared to previous

specifications without instrumentation. The coefficient on instrumented advertising is significant in three out of four reported specifications at the 28% level or below. Its magnitude ranges from -2.7 to -4.8 which is more than double the size of the coefficient when advertising is not instrumented. Nevertheless, the deterrent effect is fairly small. For example, suppose a drug had ten million dollars in revenue and one expected entrant. If the firm doubles its advertising budget from two to four million dollars, it can decrease the expected number of entrants to 0.85. At an expected entry level of two, an additional two million dollars of advertising expenditure will reduce expected entry from 2 to 1.69. Although the deterrent effect is absolutely larger, the higher number of expected entrants reduces the potential profit earned from deterrence.

The model outlined above suggests that the entry-deterring effect of advertising might only be present in a certain size market. I cannot interact advertising with certain size markets because as the number of observations interacted drops, convergence becomes difficult. Larger groupings allow convergence, but return coefficients of similar magnitude as the entire sample. In future research I plan to increase the sample size and investigate the effect of advertising on different "sizes" of drugs.

VII. Conclusion

Advertising to physicians by firms selling patented drugs discourages generic entry. However, the coefficient on instrumented advertising is neither large enough nor significant enough to substantively affect the expected number of entrants. Other characteristics of the pre-expiration market are significant predictors of generic entry.

The revenue of the brand in the year before patent expiration is the most important factor determining the amount of generic entry; higher revenues attract a greater number of entrants. The extent to which a branded drug is purchased by hospitals positively affects the number of entrants. Drugs which treat chronic conditions also attract more generic firms; the patients buying these drugs may have a higher elasticity of demand. Forms taking a long time to receive approval from the FDA experience one or two fewer entrants on average. Duopoly markets discourage generic entry as market revenue grows; the additional firm may reduce generic profits. The number of therapeutic substitutes already off-patent negatively affects the amount of generic entry, but is significant only for late entrants. Not surprisingly, the stringency of FDA regulatory procedure negatively affects the number of generic entrants observed in a market. In particular, the generic scandal of 1989 caused a significant decrease in the number of generic firms approved around patent expiration.

The previous literature has assumed that advertising before patent expiration is exogenous with respect to the amount of generic entry. The coefficient on included advertising is then negative but very small, implying that advertising to physicians does not deter generic entry very much. When instrumented in an equation predicting generic entry, the coefficient on advertising more than doubles, although its significance falls. I conclude that brand advertising can be a barrier to entry in the pharmaceutical industry, but the effect is not sufficiently strong or significant to be of concern to entrants or policy makers.

References

Bain, J. (1956) *Barriers to New Competition*, Harvard University Press, Cambridge.

Bagwell, Kyle and Ramey, Garey, (1988) "Advertising and Limit Pricing," *RAND Journal* 19:1.

Caves, R., Whinston, M., and Hurwitz, M. (1991) "Patent Expiration, Entry, and Competition in the U.S. Pharmaceutical Industry," *Brookings Papers*.

Dixit, A., (1980) "The Role of Investment in Entry-Deterrence," *Economic Journal* 90:95-106.

Dorfman, Robert and Steiner, Peter (1954) "Optimal Advertising and Optimal Quantity," *American Economic Review* 44:826-836.

Gaskins, Darius (1971) "Dynamic Limit Pricing: Optimal Pricing Under Threat of Entry," *Journal of Economic Theory* 3:306-322.

Grabowski, Henry, and Vernon, John (1992) "Brand Loyalty, Entry, and Price Competition in Pharmaceuticals after the 1984 Drug Act," *Journal of Law and Economics* 35.

Hausman, J., Hall, B., and Griliches, Z. (1984) "Econometric Models for Count Data with an Application to the Patents-R & D Relationship," *Econometrica* 52:4:909-938.

Hurwitz, M. and Caves, R. (1988) "Persuasion or Information? Promotion and the Shares of Brand Name and Generic Pharmaceuticals," *Journal of Law and Economics* 31.

Klemperer, Paul (1987), "Entry Deterrence in Markets with Consumer Switching Costs," *Economic Journal* 97:99-117.

Klemperer, Paul (1988), "Welfare Effects of Entry into Markets with Switching Costs," *The Journal of Industrial Economics* 37, December.

Leffler, Keith (1981) "Persuasion or Information? The Economics of Prescription Drug Advertising," *Journal of Law and Economics* 24:45-74.

Little, John, (1979) " Aggregate Advertising Models: The State of the Art, "*Operations Research* 27:4:629-667.

Milgrom P., and Roberts, J., (1982) "Predation, Reputation, and Entry Deterrence,"

Journal of Economic Theory:27:280-312.

Scherer, F. M., (1993) "Pricing, Profits, and Technological Progress in the Pharmaceutical Industry," *Journal of Economic Perspectives* 7:3:97-115.

Schmalensee R., (1974) "Brand Loyalty and Barriers to Entry," *Southern Economic Journal* 40:579-588.

Schmalensee R., (1982) "Product Differentiation Advantages of Pioneering Brands," *American Economic Review* 72:349-365.

Schmalensee R., (1983) "Advertising and Entry Deterrence: An Exploratory Model" *Journal of Political Economy* 91:636-653.

Scott Morton, F. (1994) "The Stragegic Response by Pharmaceutical Firms to the MFN Clause in the Medicaid Law of 1990," MIT mimeo.

Spence, Michael (1980) "Notes on Advertising, Economies of Scale, and Entry Barriers," *Quarterly Journal of Economics*:95:3:493-507.

Staiger, D., and Stock, J., (1993) "Instrumental Variables Regression with Weak Instruments," Kennedy School of Government mimeo.

Sutton, John, (1991) *Sunk Costs and Market Structure*, MIT Press, Cambridge.

Telser, Lester *et al.*, (1975) "The Theory of Supply with Application to the Ethical Pharmaceutical Industry," *Journal of Law and Economics* 18:449.

Wooldrige, Jeffrey, (1991) "On the Application of Robust, Regression-based Diagnostics to Models of Conditional Means and Conditional Variances," *Journal of Econometrics*: 45:5-46.

Table I: Distribution of Form					
	oral solid	liquid		other (cream,patch)	Total
		injectable	non-inject		
number of drug-form observations	73	28	23	18	142

Table II: Distribution of Observations among Therapeutic Categories			
Cardiovascular	20	Psychotherapeutics	24
Specific Antagonists	11	Arthritis	7
Cough & Cold	4	Anti-inflammatory	10
Dermatological	8	Allergy	13
Muscle Relaxants	6	Analgesic & Anesthetic	11
Hormones	10	Total Small Categories	18

Table III: Number of Concentrations per Drug-Form			
Number of Concentrations	injectable	other forms	oral solid
1	20	27	27
2	6	9	23
3	2	3	14
4	0	0	3
5	0	2	3
6	0	0	3
Total = 142	28	41	73

Table IV: Descriptive Statistics				
Dependent variables	Obs	Min	Max	Mean
Total	121	0	13	1.60
Middle	140	0	12	.99
Early	142	0	7	.42
Late	121	0	6	.60
Applic-Total	121	0	14	1.59
Applic-Middle	140	0	14	1.06
Applic-Early	142	0	7	.37
Applic-Late	121	0	9	.51
Independent variables and advertising				
Duopoly dummy	142	0	1	.09
Chronic disease dummy	142	0	1	.54
Injectable dummy	142	0	1	.20
Other forms dummy (cream, inhaler)	142	0	1	.29
% revenue from hospital sales (weighted by revenue)	128	0	1	.09
# of strengths per drug-form-exp	142	0	6	1.84
drug-form revenue in year before patent expiration ('000)	131	2.4	154,513	8,151
# on-patent brands in therapeutic class	139	0	12	2.20
# off-patent brands in therapeutic class	139	0	13	3.35
# labelers in therapeutic class	139	0	58	8.63
$ expenditure on journal ads 3 and 2 years before expiration	142	0	22,203	1,682
$ expenditure on journal ads 3, 2 and 1 years before expiration	142	0	32,543	2,295
# physicians who might use drug 3 years before expiration	142	2,160	254,866	136,512
months on patent	142	1	346	110
# expirations per year '86-'91	142	8	48	28.5
# drug-form-concentrations = 272; # drug-form-expirations = 142; # drugs = 113;				

Table V: Distribution of Markets by Amount of Entry at Three Moments in Time (T = patent expiration date)

Generic entrant count	# ANDAs T + 0 months = *Early*	# ANDAs T + 12 months = *Middle*	# ANDAs T + 24 months = *Total*	Applications T - 18 months = *Applic_Early*	Applications T - 6 months = *Applic_Middle*	Applications T + 6 months = *Applic_Total*
0	120	102	73	118	101	75
1	9	8	14	13	11	16
2	2	11	6	4	5	4
3	4	4	5	4	5	5
4	4	5	5	0	2	1
5	1	1	3	1	7	6
6	1	2	8	0	5	4
7	1	3	2	2	1	4
8	0	1	0	0	1	1
9	0	1	1	0	0	1
10	0	1	0	0	1	0
11	0	1	1	0	0	1
12	0	0	2	0	0	1
13	0	0	1	0	0	1
14	0	0	0	0	1	1
Obs	142	140	121	142	140	121

Table VI: Number of later entrants in markets by amount of early entry

	Early=0				Early>0			
	Mean	Std Dev	Max	Obs	Mean	Std Dev	Max	Obs
Total	.784	1.76	11	99	5.57	3.50	13	22
Late	.461	1.01	5	99	1.19	1.36	4	22

Table VII: Journal Advertising Trends[41]			
Dependent Variable	log(ads)	log(ads)	log(ads)
Time	-.0077 (.0016)	-.0071 (.0016)	-.0019 (.0028)
Postexp shift	-.149 (.051)	-.059 (.071)	-.1958 (.0933)
Postexp Time	---	-.016 (.009)	-.0213 (.0090)
Pre-exp shift	---	---	-.1395 (.0619)
Drug Dummies?	yes	yes	yes
N Adjusted R^2	6769 0.836	6769 0.836	6769 0.837

Table VIII: Group Divisions		
Revenue = Sum of hospital and drugstore revenue of drug-form in the year before patent expiration in '000 of dollars		# drugs
Group 1	Revenue < 300	29
Group 2	300 < Revenue < 2000	36
Group 3	2000 < Revenue < 8000	33
Group 4	Revenue > 8000	33

[41] Standard errors are in parentheses.

Table IX: Advertising Trend Results by Revenue Group[42]				
Dependent Variable	log(ads)	log(ads)	log(ads)	log(ads)
Time-1	-.010 (.002)	-.013 (.004)	-.011 (.004)	-.008 (.005)
Time-2	-.014 (.002)	-.013 (.004)	-.013 (.004)	-.012 (.005)
Time-3	-.006 (.003)	-.004 (.004)	-.001 (.005)	.018 (.005)
Time-4	-.0002 (.003)	-.008 (.004)	.021 (.005)	-.013 (.005)
Postexp shift-1	.112 (.091)	.245 (.126)	.315 (.163)	.129 (.162)
Postexp shift-2	.059 (.095)	.050 (.131)	.030 (.170)	.042 (.165)
Postexp shift-3	-.091 (.101)	-.175 (.141)	-.223 (.181)	-.670 (.172)
Postexp shift-4	-.778 (.130)	-1.06 (.140)	-1.02 (.182)	-.364 (.176)
Pre-exp shift-1	---	.146 (.089)	.111 (.107)	.077 (.105)
Pre-exp shift-2	---	-.003 (.097)	.031 (.112)	.009 (.111)
Pre-exp shift-3	---	-.083 (.099)	-.134 (.119)	-.386 (.116)
Pre-exp shift-4	---	-.296 (.099)	-.634 (.119)	-.153 (.117)
Postexp Time-1	---	---	-.018 (.016)	---
Postexp Time-2	---	---	.006 (.017)	---
Postexp Time-3	---	---	-.005 (.018)	---
Postexp Time-4	---	---	-.078 (.019)	---
ShareHospital*Postexp dummy	---	---	---	.031 (.078)
Duopoly*Postexp dummy	---	---	---	-.218 (.120)
Early*Postexp dummy	---	---	---	-.153 (.029)
Drug Dummies?	yes	yes	yes	yes
N Adjusted R^2	6769 0.836	6769 0.868	6769 0.839	6769 0.856

[42] Standard errors are in parentheses.

Table X: Brand Price and Quantity Trends in Hospitals and Drugstores				
Dependent Variable	log(Hospital Price)	log(Drugstore Price)	log(Hospital Quantity)	log(Drugstore Quantity)
Time	.013 (.006)	.021 (.004)	.041 (.012)	.048 (.011)
Postexp Time	-.013 (.016)	-.017 (.010)	-.049 (.029)	-.047 (.026)
Drug dummies?	yes	yes	yes	yes
N Adjusted R^2	916 0.966	944 0.983	916 0.988	944 0.995

Table XI: Brand Price and Quantity Trends by Group				
Dependent Variable	log(Hospital Price)	log(Drugstore Price)	log(Hospital Quantity)	log(Drugstore Quantity)
Time-1	.008 (.017)	-.003 (.011)	.003 (.032)	.019 (.027)
Time-2	.004 (.013)	.020 (.008)	.052 (.024)	.065 (.020)
Time-3	.022 (.012)	.031 (.008)	.062 (.022)	.037 (.021)
Time-4	.015 (.012)	.026 (.007)	.030 (.022)	.056 (.019)
Postexp Time-1	.038 (.039)	.036 (.025)	.067 (.073)	.067 (.063)
Postexp Time-2	-.009 (.031)	-.018 (.019)	-.007 (.057)	-.053 (.048)
Postexp Time-3	-.051 (.029)	-.052 (.020)	-.090 (.053)	-.036 (.051)
Postexp Time-4	-.008 (.029)	-.013 (.019)	-.113 (.055)	-.121 (.047)
Drug Dummies?	yes	yes	yes	yes
Adj R^2 N	0.966 916	0.980 944	0.989 916	0.995 944

(Standard errors are in parentheses.)

Table XII: Determinants of Generic Entry Excluding Advertising

Dependent Variable	Early	Middle	Total	Late
Ln Revenue	.651 * (.143)	.828 * (.173)	.727 * (.047)	.472 * (.079)
Duopoly*Ln Revenue	-.722 (.512)	-.722 * (.332)	-.793 * (.080)	-.439 * (.174)
Duopoly	.714 (.656)	.142 (.818)	-.039 (.390)	.557 (.541)
Share Hospital	1.03 (.621)	.800 (.745)	.922 * (.310)	1.07 * (.608)
Substitutes Off Patent	.230 (.875)	-.617 (.840)	-.389 (.307)	.230 (.579)
Other Forms	-2.06 * (.193)	-1.23 * (.425)	-.933 * (.169)	-.328 (.302)
Chronic	.640 (.422)	.556 * (.334)	.621 * (.177)	.775 (.481)
Number Concentrations	-.107 (.223)	.020 (.265)	-.0008 (.106)	-.085 (.096)
Dum87	.773 (.973)	-.061 (.892)	-.550 (.448)	-2.57 * (.193)
Dum88	1.23 (.859)	.639 (.679)	-.048 (.384)	-2.29 * (.136)
Dum89	1.03 (1.14)	-.359 (1.06)	-.636 (.424)	-1.49 * (.429)
Dum90	-1.75 * (.247)	-2.40 (2.37)	-1.19 * (.348)	-.128 (.831)
Dum91	-1.47 * (.228)	-1.65 * (.104)	-2.78 * (.113)	-1.93 * (.334)
Constant	-.964 (.666)	.827 (.808)	1.32 * (.331)	-.037 (.708)
N Pseudo R^2	128 0.368	126 0.476	110 0.414	110 0.278

(Standard errors are in parentheses. * denotes a coefficient significant at the 10% level.)

Table XIII: Determinants of Generic Entry Excluding Advertising				
Dependent Variable	Applic-Early	Applic-Middle	Applic-Total	Applic-Late
Ln Revenue	.655 * (.056)	.625 * (.042)	.740 * (.046)	.582 * (.141)
Duopoly*Ln Revenue	.126 (.588)	-.274 * (.141)	-.551 * (.120)	.071 (.399)
Duopoly	1.20 * (.400)	.473 * (.258)	-.035 (.336)	.319 (.374)
Share Hospital	-1.89 * (.172)	.778 * (.269)	.706 * (.314)	.302 (.597)
Substitutes Off Patent	.002 (.037)	-.007 (.210)	-.449 (.289)	-.591 * (.351)
Other Forms	-1.19 * (.033)	-1.10 * (.145)	-1.10 * (.152)	-1.03 * (.502)
Chronic	.463 * (.023)	.430 * (.153)	.218 (.165)	-.771 (.489)
Number of Concentrations	-.313 (.271)	-.085 (.070)	.013 (.088)	.350 * (.093)
Dum87	.865 * (.476)	-.103 (.340)	-.309 (.389)	.283 (.386)
Dum88	.078 (.290)	.652 * (.325)	.391 (.468)	.890 * (.426)
Dum89	.386 (5.55)	.128 (.434)	-.291 (.435)	-1.93 * (.776)
Dum90	.215 (4.67)	-.095 (.312)	-1.02 * (.329)	---
Dum91	-.252 (1.11)	-1.03 * (.330)	-3.27 * (.070)	---
Constant	-.206 (.449)	.560 * (.217)	1.52 * (.283)	-.107 (.546)
N Pseudo R^2	128 0.333	126 0.340	110 0.420	110 0.413

(Standard errors are in parentheses. * denotes a coefficient significant at the 10% level. Dum90 and Dum91 are omitted from *Applic-Late* because there are not enough entrants in those categories to estimate their coefficients.)

Table XIV: Marginal Effects

	Early	Middle	Total	Late	Applic -Early	Applic-Middle	Applic-Total	Applic-Late
Ln Revenue	.285	1.16	1.51	.324	.245	.750	1.44	.350
Duopoly* Ln Revenue	-.316	-1.00	1.65	-.301	-.071	-.329	1.07	.043
Duopoly	.312	.196	.081	.382	.452	.768	.068	.191
Share Hospital	.451	1.12	1.92	.734	.283	.725	1.37	.453
Substitutes Off Patent	.101	-.851	.809	.158	.021	-.008	.871	.034
Other Forms	-.900	-1.72	-1.94	-.226	-.448	-1.32	-2.14	-.717
Chronic	.280	.778	1.29	.532	.183	.517	.423	-.462
# Concen-trations	-.047	.028	-.002	-.059	-.082	-.102	.026	-.21
Dum87	.338	-.085	1.14	1.77	.142	-.124	-.599	.170
Dum88	.538	.882	.100	1.58	.160	.782	.759	.534
Dum89	.452	-.495	1.32	1.03	.082	.154	-.398	-1.16
Dum90	-.765	-3.32	2.48	.088	.168	-.114	-1.97	---
Dum91	-.641	-2.24	5.78	1.34	-.078	-1.24	-6.35	---

Table XV: Determinants of Generic Entry with Advertising Included as an Exogenous Variable

Dependent Variable	Middle	Total	Applic_Middle	Applic_Total
Advertising	-1.56 (.530)	-.252 (.663)	-.462 (.133)	-1.51 (.665)
Ln Revenue	1.07 (.129)	.739 (.082)	.771 (.104)	.865 (.092)
Duopoly*Ln Revenue	-.713 (.168)	-.682 (.119)	-.143 (.159)	-.324 (.164)
Duopoly	.630 (.244)	.108 (.260)	.784 (.243)	.560 (.246)
Share Hospital	1.09 (.480)	.731 (.315)	1.14 (.401)	.959 (.343)
Substitutes Off Patent	-.153 (.222)	-.010 (.172)	.590 (.220)	.103 (.183)
Other Forms	-1.07 (.355)	-.999 (.256)	-1.46 (.351)	-1.32 (.296)
Chronic	1.01 (.367)	.727 (.244)	.592 (.296)	.564 (.256)
Dum87	-.068 (.307)	-.349 (.269)	-.120 (.354)	-.313 (.271)
Dum88	.981 (.242)	.335 (.224)	.918 (.269)	.548 (.222)
Dum89	-.186 (.273)	-.577 (.238)	.424 (.243)	-.197 (.228)
Dum90	-2.37 (.597)	-1.14 (.270)	-.058 (.283)	-.828 (.270)
Dum91	-1.65 (.385)	-2.56 (.723)	-1.04 (.349)	-2.91 (1.01)
Constant	.497 (.431)	1.11 (.287)	-.054 (.370)	-1.16 (.313)
N R^2	130 0.470	111 0.443	127 .419	111 0.489

(Standard errors are in parentheses.)

Table XVI: Determinants of Generic Entry (ommitting 1991) with Advertising Included as an Endogenous Variable				
Dependent Variable	Middle	Total	Applic_Middle	Applic_Total
Advertising	-3.62 (2.35)	-4.81 (4.49)	-2.69 (3.39)	-4.50 (3.69)
Ln Revenue	1.30 (.698)	.765 (.343)	.828 (.459)	.869 (.407)
Duopoly*Ln Revenue	-1.01 (.807)	-.439 (.470)	-.348 (.625)	-.193 (.602)
Duopoly	.751 (.503)	1.12 (.870)	1.03 (.719)	1.27 (.782)
Share Hospital	.648 (1.34)	.419 (.822)	.831 (1.06)	.439 (.937)
Substitutes Off Patent	-.562 (.431)	-.257 (.351)	.156 (.469)	-.208 (.364)
Other Forms	-1.41 (1.06)	-1.04 (.835)	-1.44 (1.11)	-1.54 (1.10)
Chronic	1.08 (1.14)	1.13 (.747)	.787 (.900)	.737 (.801)
Dum87	-.406 (.616)	-.591 (.564)	-.173 (.698)	-.355 (.551)
Dum88	.917 (.316)	.414 (.366)	.847 (.411)	-.641 (.344)
Dum89	-.033 (.778)	-.559 (.681)	.568 (.612)	-.102 (.597)
Dum90	-2.76 (2.57)	-.796 (.664)	-.072 (.606)	-.588 (.678)
Dum91	ommitted	ommitted	ommitted	ommitted
Constant	1.20 (1.22)	1.48 (.802)	.376 (.924)	1.60 (.867)
N R^2	97 0.679	93 0.427	97 0.471	93 0.548

(Standard errors are in parentheses.)

Data Appendix

To form the dataset I aggregated the quantity and revenue figures for different presentations of the same concentration-form of a drug. Units are comparable because the chemical, concentration, and form are consistent. I did not use IMS's data on generic entry, instead keeping only observations on the brand's revenue and quantity.

Through a Freedom of Information Act request, I received a complete list of patents monitored by the FDA. The list begins with patents expiring in 1986 and continues into the next century. First, I selected those patents expiring between 1986 and 1991. The FDA will give a firm the right to exclusively label a product if it showed the drug was appropriate for an additional "indication" or "use." I eliminated indication and use exclusivity rights because these rights only affect the labelling of the product and not which firms are allowed to manufacture it. One exception to that rule is the drug atenolol in my dataset (The brand name is "tenormin."). It had no approved indications other than those expiring after the patent on the compound expired. The indication patents were therefore binding. Additionally, doctors are aware that a particular compound is effective in treating a set of symptoms or diseases, even if labels differ across bottles.

I also eliminated patents for which there was substantial entry well before the patent indicated it was permitted; the patent had clearly been broken. Potential bias was introduced here as I could not identify non-binding patents in markets where no firm was interested in entering. Only five instances of this occurred out of an initial list of about 200 drugs, so I assume the effect is negligible. This problem is unavoidable when using pharmaceutical data; Caves, Whinston, and Hurwitz employ the same procedure. In cases where the generic entrants appeared to be late, I confirmed all patent expirations with FDA officials.

Firms do not have to specialize in either innovation or imitation. For instance, some branded firms also make generic versions of other firms' drugs, and a generic firm can get a process patent if it invents a better way of making an innovator's drug. These drugs are known as "branded generics" and may be regarded as better substitutes than

regular generics. I do not pursue that distinction here. A generic firm may try to establish a reputation by prominently displaying its name on the label of its product, or it can sell its entire output to a "labeler." A labeler is a firm that registers itself with the Health Care Financing Administration as a pharmaceutical packager.[43] A manufacturer is allowed to label all or part of its own output itself and/or to sell the drug to other labelers who will apply their name and number to the bottles. Manufacturers are confused with labellers in many data sources.

IMS contracts with a panel of 1,000 doctors to record every detailing visit and the number of minutes spent talking about each drug. A doctor may remain on the panel for as long as he or she likes. The number of minutes reported by the doctors is translated into dollar expenditure by using a constant "cost per call" rate across firms. This procedure could easily lead to imprecise estimates of detailing expenditures.

[43] The labeler is legally responsible for Medicaid rebates on that product. See Scott Morton (1994) for an analysis of the Medicaid Rebate Program.

One Possible Model of Generic Entry and Brand Advertising and Price Choice

Agency problems are the main feature of the prescription pharmaceutical market. Consumers do not have the necessary information to choose among medicines, and physicians do not pay for the drugs and so have little or no incentive to minimize cost.[44] Hospitals, HMOs and mail order houses have the correct incentive but still do not make up the majority of the market. It seems uncontroversial that a firm making a branded product facing patent expiration would like to influence the post-patent game in its favor, using whatever instrument is most effective. Below is a model including what I believe are the critical features of the market.

This model describes the choice of the physician based on his or her preferences. The physician has knowledge of the branded product due to advertising during the brand's exclusivity period. In contrast, the physician knows very little about the generic alternatives because, roughly speaking, generic firms do not advertise. The cost to the physician of prescribing the generic, a relatively unknown drug, can be categorized as follows:

(1) Medical costs of switching a medicated patient to a drug that may be metabolized slightly differently. Such a patient may be harmed by the switch to a generic alternative.

(2) Medical/Social costs of changing medication of old or easily confused patients. It may be worthwhile keeping the patient on the same drug if he or she refuses to take a pill of a different color.[45] These two explanations only apply to products treating chronic diseases.

(3) The penalty for making a low quality product is much higher for a branded manufacturer whose reputation is at risk than for a generic with less name recognition. A physician may feel that there is less risk involved in taking the branded product; its quality will be higher on average and the probability of a

---

[44] "Third-party reimbursement plans operated by the government and private insurers have expanded to cover an estimated 44 percent of prescription drug outlays in 1987, up from 28 percent in 1977 (U.S. Office of Technology Assessment, 1993, p27.)." Scherer (1993) p99.

[45] Pill shapes and colors are copyrighted by brands. A generic entrant may not make its product an exact copy of the original brand without being sued, suggesting branded firms are aware of cost (2).

defective dose will be lower.[46] In addition, the physician may be afraid of being sued if a generic drug harms a patient.

In the face of these costs, the question might be better put, why should any physician prescribe a generic drug? My assumption is that the physician gets some feedback from patients about the prices of drugs and also learns from the hospital or managed care unit he or she has contact with.

Most pharmaceutical advertising is undertaken by firms selling branded products and is directed at physicians. We can think of two types of advertising to physicians. One is disseminating information about new products; early in a product's life this type of advertising is predominant. Detail staff and journal ads explain the properties of a new drug to physicians. Later in the product's lifecycle advertisements in journals and detailing serve to remind physicians of the existence of a company and its product. Leffler finds, interestingly, that for a given physician, a disproportionate amount of detailing time is spent on products that were released when that physician was in medical school.[47] Detailers appear to be focussing on drugs physicians are most familiar with already, suggesting the brand firm wants to emphasize its quality rather than novelty.

I propose that the role of advertising late in the product lifecycle is to inform physicians about the firm's true "type," or quality. Once that physician is informed, he or she is aware of the higher quality and lower risk associated with the branded product compared to the generic. Let firm types, τ, be distributed uniformly from $\underline{\tau}$ to $\bar{\tau}$. Advertising technology is concave in the probability of reaching a given physician as expenditures rise, and to advertise at all requires a fixed cost.[48] The ex-post probability that a firm will reveal its type to a physician thus depends on advertising expenditures in a positive way.

---

[46] The generic scandal provided evidence that the FDA cannot perfectly monitor the quality of generic drugs.

[47] Leffler (1981).

[48] This general shape is suggested by Leffler (1981) and Little (1979).

Once the physician receives information about a firm, he or she has an estimate of the quality difference. The expected quality of a firm the physician has never heard of is a weighted average of low quality firms that are not advertising and the probability of not receiving information from a firm that is advertising. Advertising therefore increases a physician's cost of prescribing the generic; he or she realizes the brand is high quality. Equilibrium firm behavior is for low quality firms not to advertise; their true types would be revealed. Increasingly high quality firms advertise more. As type increases, informing the physician of the true type of the firm becomes more valuable. The cost of disseminating information more thoroughly is rising also; a firm chooses the point where marginal benefit equals marginal cost.

In turn, each physician has a type θ which is a function of his or her risk aversion and patient mix. θ is defined to fall with risk aversion. For example, a physician with many elderly patients, or patients in critical condition, might prescribe generics less often than other physicians. A physician's cost of prescribing a generic can be expressed,

$$\begin{aligned} \textit{Switching Cost} &= S(\theta, A) \\ &= g + [\min(0, A - a)]^{\beta} - \gamma\theta \end{aligned} \tag{3}$$

$$\frac{\delta S}{\delta A} > 0, \quad \frac{\delta S}{\delta \theta} < 0 \tag{4}$$

where A is the level of advertising. I choose the simplest case to analyze, a linear form for the switching costs. g is simply an intercept, and A represents advertising expenditure. I assume $0 < \beta < 1$ to reflect diminishing returns from advertising; $a > 0$ is the fixed cost of advertising. I assume the physician has some access to information about prices and availability of generics; trade publications will have studies or news about generics, pharmacists will telephone to discuss substitution, and the hospital or HMO a physician is associated with will dispense generics. The physician will prescribe a generic when,

$$P^b > P^g + g + (A-a)^{\beta} - \gamma\theta \tag{5}$$

Some physicians may have such a high type that they never prescribe generics; others may always do so. The idea is that price plays a role at the margin. In addition, there is a regulatory aspect to the choice of brand or generic. Substitution laws in the United States evolved considerably over the 1980s. Nearly all states repealed antisubstitution laws in the years before the Waxman-Hatch Act was passed (1984). Mandatory substitution laws existed in twelve states by 1993. The outcome of brand/generic choice can be shown diagrammatically for different regimes.

Prescribed	Substitution	Regime	Outcome
BRAND	no	---	BRAND
	yes	substitution forbidden	BRAND
		substitution permitted	pharmacy choice: BRAND or GENERIC
		substitution mandated	GENERIC
GENERIC	---	---	GENERIC

In the case where the pharmacy has a choice about which type of product to dispense, I make a simplifying assumption. The pharmacist telephones the doctor, whereupon the doctor decides according to the rule above. Hospitals and HMOs make up a fraction λ of the market; they are the upper end of the theta distribution and the lower end of the demand curve. I assume that their switching costs are unaffected by advertising. The buyers for large organizations are much less affected by advertising; they gather more information about quality and are less risk averse. Thus, for this group, switching costs are lower and agency problems fewer, making purchase of generics more likely.

The model has two periods.[49] Advertising level is chosen by the brand in period one. I assume the total quantity of the drug sold is fixed in the two periods. The drug is at the end of its patent life; there is unlikely to be learning and diffusion still going on. Customers are the physicians described above who have a distribution of types, θ. θ is distributed uniformly from 0 to 1. Total market quantity, Q, is normalized to 1. Entry

[49] The following analysis is based loosely on Klemperer (1987).

is permitted in period two; if entry occurs, it is undertaken by competitive generic firm(s) which compete Cournot using the residual demand curve. All firms have the same marginal cost, c. Generic firms must pay a fixed cost, F, to enter. I assume the brand's first period price, P_1^b, is given exogenously; it is determined by prices of alternative therapies.

The brand faces the standard monopoly tradeoff between raising price and losing marginal customers to the generic firm(s). The brand will maximize period 2 profits,

$$\Pi_2^b = (P_b - c)(\frac{1}{\gamma})(P_g + g + (A-a)^\beta - P_b) \tag{6}$$

by choosing a type θ^* (which is a function of the generic price), and selling only to types above it. Physician types below θ^* get more surplus from buying the generic.

$$\theta^* = \frac{1}{2\gamma}(P_g + g + (A-a)^\beta - c) \tag{7}$$

As advertising increases, optimal θ^* rises; the brand sells more. Figure 1 illustrates the linear demand curve example. Generic firms are left with the remainder of the market: $Q-q_b$. The generics trade off a higher price against losing their marginal customers to the brand; since total quantity is fixed, it is in this sense that generic firms have a downward-sloping demand curve. The residual demand curve for the generic product can be derived simply by using the expressions in equation six.

$$P_g = 2\gamma Q - g - (A-a)^\beta + c - 2\gamma q_g \tag{8}$$

Since this demand curve is linear, the symmetric Cournot competition solutions, price, quantity, and profits, are particularly straightforward to derive and I will not reproduce them here. The generic enters in period 2 if its expected profits are greater than zero. The equilibrium number of firms will be,

$$N^* = [\frac{(2\gamma Q - g\ (A-a)^\beta)^2}{2F\gamma}]^{1/2} - 1 \tag{9}$$

The brand can choose a level of advertising which maximizes its total profits, taking into account the effect of advertising on the amount of generic entry, generic price and therefore brand price. If the market is sufficiently large that an entrant can cover its fixed cost by selling to customers with low switching costs (hospitals and HMOs), it will be impossible for the brand to deter all entry. Alternatively, if the market is so small that total profits would not cover the fixed cost of entry, the issue of deterrence is moot. In between these extremes is a region where the brand can affect the post-patent game by its choice of advertising level in period one. This option for the brand can be most clearly seen in Figure 2. Advertising essentially lifts the demand curve of the brand, while leaving the lower end (hospitals and HMOs) unchanged. This creates a discontinuity in the brand's profit function which more easily allows it to choose an effective, entry-deterring price.

Switching Costs and Brand Market Share Figure 1

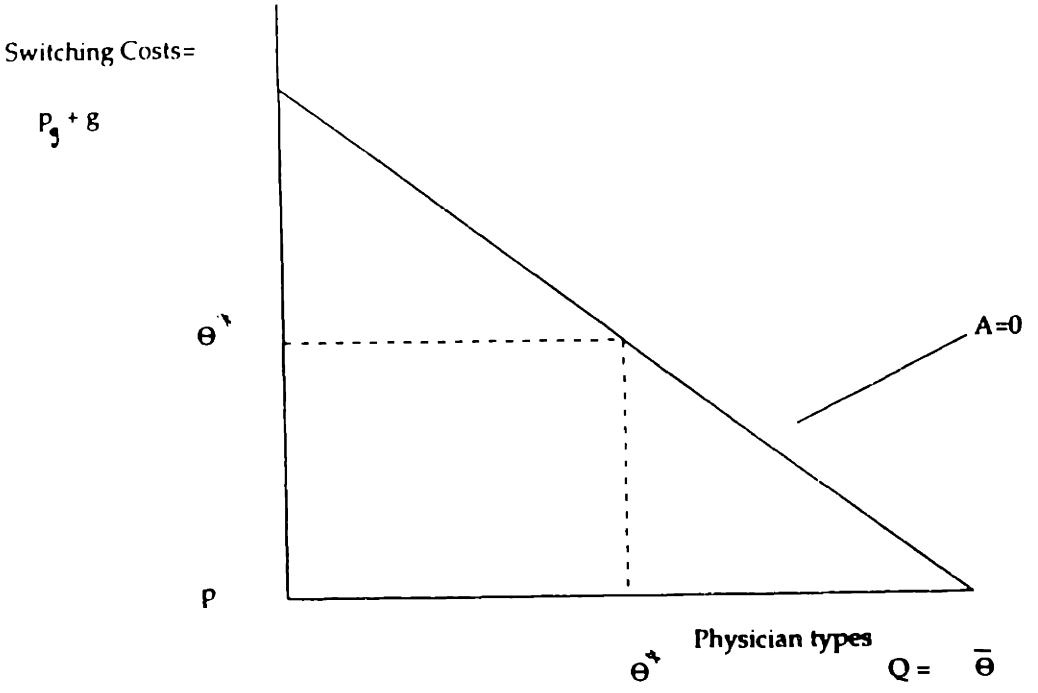
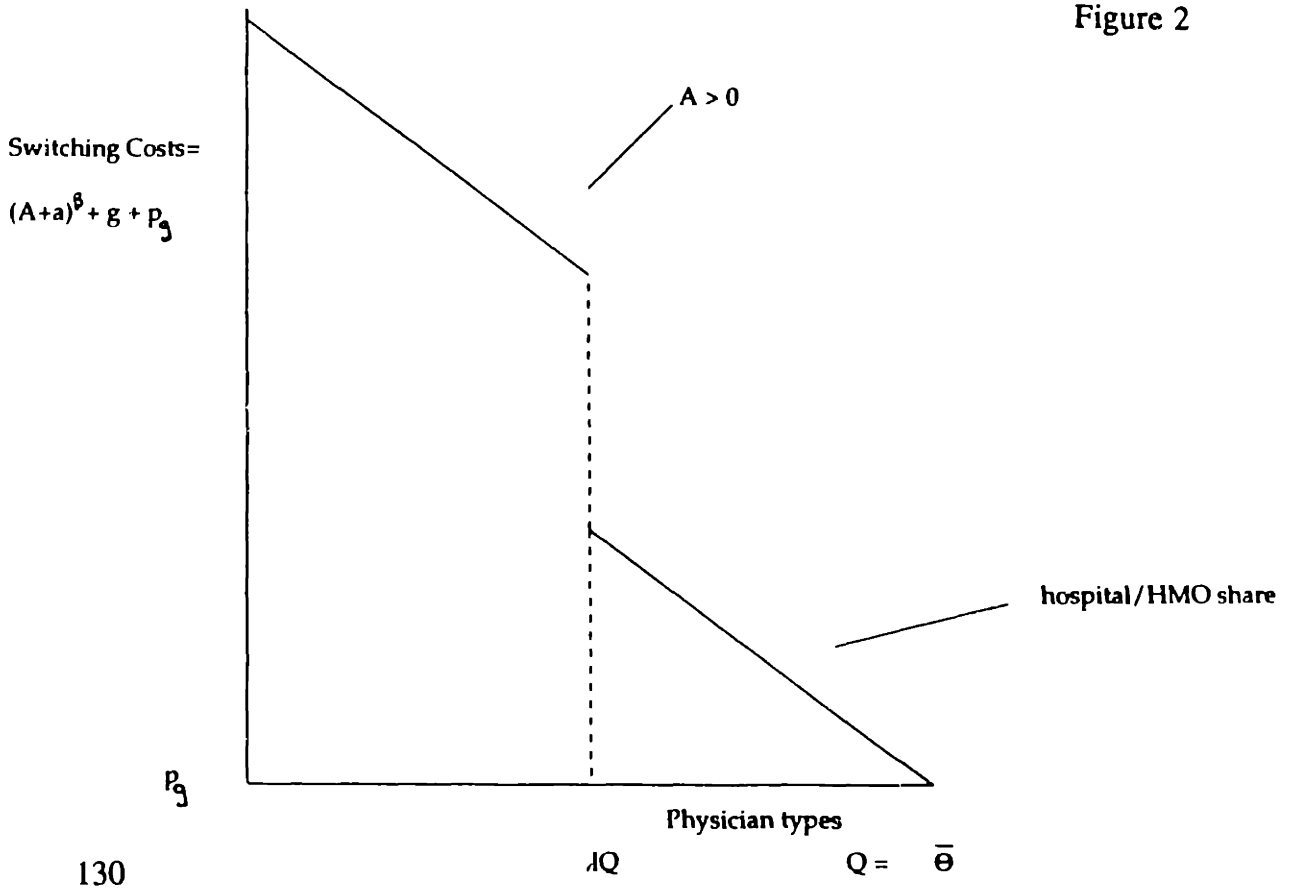


Figure 2



Chapter III

The Strategic Response by Pharmaceutical Firms to the MFN Clause in the Medicaid Rebate Law of 1990

Introduction

Most Favored Nation clauses have been studied extensively in the Industrial Organization theoretical literature; see, for example, papers by Cooper (1986 and 1991), Png (1987 and 1991), and Salop (1986). The basic model of MFN clauses has the following features. A firm announces it will offer its lowest price to a specific set of customers.[50] The firm will then find that competing with other firms for heterogeneous consumers on the basis of price becomes costly. Any price discount given to a marginal customer to induce a sale must be applied to all customers covered by the MFN. Theory tells us that firms that credibly adopt the MFN clause can commit to "soft" price competition. Although the literature contains many models explaining the strategic effect of MFN clauses, there has been relatively little empirical verification of the effectiveness of the policy. The Medicaid rebate program which began in January 1991 included an MFN provision for Medicaid reimbursement; Medicaid purchases are roughly fifteen percent of the prescription pharmaceutical market. This paper examines the perhaps-unforeseen strategic effect of the policy on average pharmaceutical prices. I examine how the Medicaid Rebate law changed prices and quantities sold by firms in different competitive positions. I briefly discuss the welfare implications of the law and ask whether the law was overall a good policy choice.

I find that after the Medicaid rebate law took effect, the average price of branded drugs facing generic competition increased from about three to five percent. The average

[50]The firm chooses the applicable time period; it could extend the guarantee to past, future, or both sets of customers.

price of a patented drug rose a little more, from about four to nine percent, because of the MFN clause. I do not find good evidence of a price rise among generic manufacturers facing MFN-constrained brand competitors. In general however, the MFN clause forced producers to engage in soft price competition; the producers responded by raising their prices, as theory would predict. Although the Federal and state governments saved 150 million dollars per quarter in Medicaid expenditure by the end of the first year of the program, average market prices rose as a consequence of the reimbursement policy. Other government expenditure, such as purchases made by the V.A., increased due to these rising prices and partially offset the initial gain. Non-Medicaid consumers of pharmaceuticals, especially those who obtained their drugs at retail pharmacies, paid higher prices also.

The structure of this paper is as follows. The exact rules of the rebate scheme are explained in Section I. Section II discusses several theories of price discrimination and the effects of an MFN clause within and across firms. The pharmaceutical industry, its rules of entry, and the available data are discussed in Section III. Section IV contains an analysis of the patterns of observable price dispersion. The estimation of the effects of this MFN clause on price and quantities is reported in Section V. Section VI concludes.

I. The Medicaid Rebate Law

The Omnibus Budget Reconciliation Act of October 1990 (OBRA 90) limited Medicaid reimbursements for drug purchases to a fraction of the average wholesale price

for each drug. The idea behind the law was that Medicaid was not receiving the low prices given to other big buyers because it reimbursed individual hospitals and pharmacies rather than purchasing in bulk. Pharmaceutical firms engage in a great deal of third-degree price discrimination; they do not use the uniform price so common in theoretical literature. The fact that Medicaid could not use its bargaining power to secure the normal advantages of a large buyer in a market with significant price dispersion was part of the impetus for the legislation.

To secure better prices for Medicaid (and reduce government expenditure) OBRA 90 required drug manufacturers to pay rebates to state and federal Medicaid programs. A rebate would represent the total dollar amount by which a firm had been overpaid by Medicaid in that quarter, compared to the new low prices that were now required. Manufacturers of branded pharmaceuticals were required to sell to Medicaid at 88 percent of Average Wholesale Price (AWP) or their "best price," whichever was lower. Average Wholesale Price is simply a quantity weighted average of a firm's wholesale prices. Thus if a firm sold one unit of its product to a customer at 60 percent of AWP, it would effectively have to sell all its Medicaid units at that price also (although the mechanism would be a rebate check). Generic products were not subject to the same scheme; instead a generic product's price to Medicaid was required to be 90 percent of its AWP.[51] The fact that generic products are not subject to any "best price" provision turns out to be an important distinction.

---

[51] AWP is a quantity weighted arithmetic average that includes all package sizes of a given drug sold to pharmacies.

The law required that the price used to calculate "best prices" and rebates be the *price per unit* of the drug; common units are the pill or the milliliter. Separate packaging alone would not constitute a new product with a new "best price." Rather, the firm would have to calculate the price per pill on all its packages of a given drug to find the lowest price. Additionally, OBRA 90 defined best price to be the lowest contract or invoice price available to a member of the "retail pharmaceutical trade."[52] Only prices of goods sold to drug stores or HMO's, not hospitals, counted in the calculation of "best price." Thus, a firm would lose revenue on all its Medicaid sales each time it gave a *pharmacy or HMO* a price lower than 88 perc of its AWP. The exact amount owed by each firm was calculated every quarter, using sales data from that quarter, by the Office of the Inspector General. (The data are highly confidential; only the Health Care Financing Administration (HCFA) sees the figures.) Drug manufacturers rebated $150 million per quarter to state and Federal Medicaid programs at the end of 1991.

II Theory of Price Discrimination and Most Favored Nation Clauses

How would we expect price and quantity to move after a law such as the Medicaid rebate law is enacted? Both generic and patented brands are facing a mandatory discount on all Medicaid sales, on average 15 percent of sales. If the discount

[52] Cash rebates are common in pharmaceutical purchasing. For example, a customer might get 2% of the invoice total returned as a rebate. The customer has to earn the rebate by paying in cash, for example, or increasing usage of the drug within their HMO. Firms are explicitly instructed to incorporate information on cash rebates into their calculation of "best price." Therefore, firms have no particular incentive to alter the size or composition of rebates.

price is above marginal cost, the firm would rather sell at that price than refuse to sell, so the government is in a good bargaining position.[53] Using the notation of Cooper (1986) (who actually uses a two-period model) we see that the critical feature of an MFN clause is that the profit function of the firm is altered. Instead of profit Π, the firm faces,

$$\Pi_{MFN} = \Pi_{Original} - (p_i - p_j)Q_i \quad if \quad p_j < p_i \tag{10}$$

where p_i is price for customer i who has just bought amount Q_i, and j is another consumer who is offered price p_j. When the firm maximizes the new profit function with respect to customer j's price, it will have an additional term that is positive in p_j. The firm will earn a penalty if it sets $p_j < p_i$; in general it will earn higher profits if it sets a higher price. The important strategic effect, which Cooper goes on to show, is that even a firm that *unilaterally* institutes MFN pricing alters its own best response function so that Bertrand equilibrium prices for both firms are higher.[54]

Oligopolistic firms that are competing for the same customers, as therapeutic substitutes do, will become less aggressive in price competition under an MFN, causing market prices to rise. Quantity sold will decline according to the elasticity of demand for the products; for pharmaceuticals, the elasticity of demand could be quite low, which would imply that a price rise would lead to increased industry revenues. In the simple

---

[53] Firms may decline to have a product participate in the Medicaid Rebate Program (i.e. not pay any rebates on that drug to the states). A drug which is not participating may be excluded from a state's Medicaid formulary; its purchase will not be reimbursed. Any drug which participates in the Rebate Program may not be excluded from any state's formulary.

[54] Cooper 1986.

framework described thus far, the firm's uniform price simply rises. However, the pharmaceutical firms in my sample are not selling at a uniform price; rather each has a pre-existing distribution of prices.

When an MFN clause is implemented, a firm which price discriminates has a motivation to raise the prices at the bottom of its price distribution. Therefore, the degree of price discrimination employed by different types of firms (such as brand v generic or monopoly v competitive firms) is important for analyzing the results of the MFN. Later sections of the paper look at the extent of price discrimination for different types of firms before and after the policy change. Therefore, there are two effects working to raise prices of pharmaceutical products: the within-firm effect caused by price dispersion and the well-known strategic effect. Some categories of firms may have more price dispersion than others, and so will find the within-firm effect stronger.

A. Within-Firm Effect

The baseline case of uniform pricing is helpful to think about first. Under a uniform price rule the firm offers the same unit price to all customers, regardless of the quantity bought. A best price rule will have no effect on firm pricing or net prices because the difference between the minimum price and the price paid by consumer i, Medicaid perhaps, is zero. However, a rule which mandates a discount to some customers based on average price will alter the firm's profit function. The equilibrium price with the rebate scheme will be higher than without it. In general, there are two basic effects we expect to see *within the firm* after the OBRA 90 legislation took effect:

a response to the "best price" provision and a response to the "average price" rule. The behavior of any one firm may not exhibit both, depending on its pre-existing pricing scheme - but both are caused by the existence of price dispersion.

The existence and form of price discrimination within a firm can be explained by a number of theoretical models in the mechanism design area. The firm's optimal response to the government's rebate scheme will vary across models. For example, perfect price discrimination and second degree price discrimination are much more complex situations than that of uniform pricing. The distribution of consumer valuations and the valuation function of the consumers will determine both the firm's original pricing scheme and its optimal response to the legislation.[55]

B. Between-Firm Effects

An additional complication is that the firm analyzed thus far may not be a monopolist. If the brand producer faces generic competition, it will be in one of two positions. If it has kept some high-elasticity customers by price discriminating, it will likely lose them to generic firms if it stops price discriminating. Alternatively, the brand may only be serving low-elasticity customers and not price discriminating as much. If the brand is protected by a patent, the firm may price discriminate to compete for low valuation consumers who might buy therapeutic substitutes. When prices rise due to the MFN law, some former low valuation customers may no longer demand the product, or demand less of it. If a brand had no substitute, it would not alter its price or quantity in response to between-firm effects. However, my dataset consists entirely of drugs

[55] This area is the subject of ongoing research by myself and Peter Klibanoff.

classified by IMS to be "cardiovasculars;" many of the drugs are therapeutic substitutes for one another.

Generic firms will have an incentive to raise their prices slightly because of the average price provision they are subject to. However, the quantity and price responses of generic firms will also depend on how many generic firms are in the market with the competitive brand. When one firm adopts an MFN clause its rival has an incentive to raise its price. If the MFN firm has many rivals, all producing a homogeneous product and competing in price, each of the rivals has much less incentive to give a soft response; they are all still competing with each other. Since only the brand has shifted its best response curve, there may be no price rise by generic rivals.[56] In the case of generic pharmaceutical firms, the market is a good approximation to Bertrand competition with perfect substitutes, so the changed brand behavior will have no effect on generic prices if there is more than one generic firm in the market. On the other hand, if the generic is the sole competitor, it will raise its prices in response to the softer competition caused by the MFN. Some of the brand's customers have sufficiently elastic demand that they are willing to switch to a generic when the brand price rises. Sole-competitor generic firms should experience an increase in quantity as well as price.[57]

Usually when industry prices and profits rise we expect to see entry and more vigorous competition. Patented drugs are obviously protected from entry by their patent.

---

[56] See Cooper (1986) for a theoretical treatment.

[57] One explanation for a price rise in the case of single generics might be shortage of capacity. This explanation should not be the whole story for the following reason. A generic firm manufactures many products using mostly fungible equipment; output of any one drug is therefore not capacity constrained.

Generic manufacturers are, in general, not protected from entry. However, a generic entrant must receive approval from the FDA before it can begin selling a generic drug. The approval process takes 18 months, on average, after the application is submitted. The firm would normally spend several months preparing the application. Thus, the earliest an entrant could appear in these markets would be towards the end of 1992, well outside my sample period.

HCFA has prepared but not issued[58] a report which documents the effect of the "best price" portion of the Medicaid rebate scheme as opposed to the "average price" part. HCFA concludes that the "best price" rule is not the dominant source of rebate revenue. Their analysis however, looks at firm behavior after the MFN legislation is passed. The reason the best price provision does not directly trigger many rebates could be because firms have raised their lowest prices until the best price rule does not bind. The evidence in the section above supports the hypothesis that firms reduced the amount of dispersion in their price distributions after the MFN law was passed.

Recall that the Medicaid Rebate legislation defines the "best price" for purposes of the rebates to be the lowest price the firm offers to the "retail line of trade." The retail line of trade excludes hospitals and thus, price discounts given to hospitals do not affect the calculation of either AWP or best price. A firm will have no direct reason to change its pricing policy in the hospital sector. When examining price and quantity changes in the drugstore sector, one can compare them to changes in the hospital sector; the hospital sector changes should control for unexpected idiosyncratic price and quantity

[58] At the time of writing, Congress had not released the document.

movements of the drugs in my sample. The fact that the hospital sector should not react to the policy change allows a difference analysis to be carried out.

However, the difference results should be interpreted cautiously. The hospital sector's price schedule may be indirectly affected by pricing policies in the drugstore sector by firms worried about arbitrage, for example.[59] Additionally, hospitals are known to receive more cash discounts than pharmacies. Those cash discounts do not appear on the invoices that comprise the dataset. If invoice prices change and cash discounts compensate, the data will show a change when none has occurred. However, the Rebate Law explicitly instructed firms to include cash discounts in their calculation of best and average prices; thus the law did not provide any incentive for firms to alter their pricing policies. Nevertheless, the levels of the estimated changes in both sectors should be considered as well as the differences.

A further problem which may dilute the effects observed in the data is that these data do not contain purchases made by HMOs. HMO prices and quantities are reported separately by IMS. To the extent that the binding low price in 1991 was a price to an HMO, drugstore prices may not be altered. Changes to prices in the drugstore sector will be rearrangements of the price schedule that do not necessarily change the mean. Below in Table I are listed the expected effects on pre-rebate average price, given that hospital prices and quantities do not respond to the law, but drugstore prices do. The

---

[59] A pharmaceutical manufacturer might be worried about arbitrage between hospitals and pharmacies, or more generally between low and high valuation customers. If arbitrage were easy and prevalent, price discrimination would be much less effective, or perhaps impossible. There are almost certainly some leakages out of hospital pharmacies, but I believe the amounts are not significant.

table shows the change in quantities and prices, before rebates are paid, for each category of drug and each retail sector at the imposition of an MFN law. The pre-rebate mean price is exactly what the IMS data record; the prices in the data should respond to the imposition of the MFN clause according to the strength of the strategic effect and the extent of brand price dispersion in the drugstore sector. The expected "difference in difference" results, given that hospitals are unaffected by the law, are exactly the same as the expected drugstore results here.

Table I: Expected Changes in Price and Quantity after MFN Law by Competitive Class and Distribution Channel

	Δ P		Δ Q	
	Drugstore	Hospital	Drugstore	Hospital
patented brand	+	0 ?	0/-	0 ?
competitive brand	+	0 ?	-	0 ?
off-patent brand	+	0 ?	0/-	0 ?
generic	0	0 ?	0	0 ?

From the firm's point of view, the most important question might be how much (and in what direction) post-rebate revenues change. To find the post-rebate mean price, the researcher must know where in the distribution the Medicaid purchases fall, and how many of them there are. If Medicaid sales are already at the low end of the price distribution, the direct payments required by the scheme will not lower firm revenue by very much. In general, the direct and strategic effects oppose one another and their relative strengths cannot be precisely evaluated without individual invoice data from IMS and individual Medicaid purchase data that is collected by HCFA.

III. The Pharmaceutical Industry

Manufacturers of pharmaceuticals fall into two main categories. The first category contains "innovator" firms; they undertake research and development to discover new drugs and bring them to market. Innovator firms spend vast amounts on research and development in an attempt to discover new and profitable drugs. Once approved by the FDA, such drugs are marketed under a proprietary, or brand, name by the innovator. Brand and monopoly are not synonymous. If a brand is a monopoly it falls into one of two categories; the firm might still own a binding patent or the patent may have expired but no generic has entered that market.

A second type of firm is a generic or imitator firm. After patent expiration, any firm may submit an Abbreviated New Drug Application, or ANDA, to the FDA. The generic firm must show its product is bioequivalent to the original branded product; relatively minor clinical tests are required in comparison to those required for a brand. Once its ANDA is approved and the patent has expired, a generic firm may legally make and sell the product. Usually, generic firms sell their products under the chemical (generic) name of the drug, rather than a proprietary name.[60] The ANDA approval process takes, on average, eighteen months from first submission of the application to final granting of approval.[61] Thus, some drugs have two categories of manufacturers,

---

[60] Sometimes generic firms invent their own proprietary names instead of marketing the drug under its chemical name.

[61] After it receives approval the entrant has permission to make the drug but is not obligated to. If a firm does not manufacture the drug, it does not lose the ANDA. The FDA retains discretionary control over all ANDAs and can withdraw an ANDA for inspection failures or other breaches of the regulations.

the brand and one or more generics.

The market for pharmaceuticals thus displays a wide variety of competitive conditions. At one extreme is a new branded drug that is produced under patent protection by one manufacturer, at the other is a market with many generic manufacturers. I assign all drug-manufacturer combinations to a competitive class for ease of discussion. All pioneer drugs that have patent protection are in the "patented brand" class.[62] A patented brand continually faces competition from therapeutic substitutes, despite its patent protection. Some brands are still monopolies but have lost patent protection; no generic has entered that particular market.[63] These drugs are called "off-patent brands." The level of competition in the case of a market with brand and an expired patent can be characterized by the number of ANDA's (*Number of ANDAs*) granted for a concentration and form of a drug at any given time; *Number of ANDAs* ranges from 0 to 23 in the dataset. I refer to a brand in a market with one or more generic firms as a "competitive brand." The final two classes include all the generic products: ANDA holder and generic distributor.

Price dispersion comes from two logically distinct sources. Precisely the same good may be sold at different prices to different customers. The seller might base the price on evidence that the buyer belongs to a specific group (e.g. an HMO) or on

---

[62] Sometimes two firms discover a drug independently and share the patent. The duopoly observations I have must be firms that failed to agree on a contract that would allow them to earn monopoly profits. Although the market structure is a duopoly rather than monopoly, I class these observations with "patented brands."

[63] See Scott Morton (1993) for a discussion of why an expired patent might not attract entry from generic firms.

knowledge of the buyer's elasticity of demand. Except for the distinction between drugstore and hospital sales, this source of price dispersion is unobservable in my dataset; IMS reports an average price. However, the other source of dispersion is a type of quantity discounting, which I do observe. For example, tablets can come in bottles of 10 or 1,000. The latter might not be very practical for a small pharmacy although the cost per pill is usually lower. Manufacturers also use special types of packaging to take advantage of heterogeneous consumers. Proprietary convenience packaging like "accudose" packs that mark a patient's daily dose, or vials that come with special types of IV equipment, contain the same chemical entity as simpler presentations but cost much more.

Manufacturers or wholesalers may legally offer a different (always lower) list price to non-profit hospitals than to pharmacies because the Robinson Patman Act exempts non-profit buyers such as hospitals from its prohibitions against price discrimination. Hospitals and HMOs can get contracts from manufacturers that guarantee a fixed price for one or two years; they are large enough buyers that they can demand some price stability. Small pharmacies are not usually offered contracts. A well-known characteristic of the market is the existence of cash rebates. Rebates are simply cash returned to the customer after buying a certain quantity of a drug, a particular mix of drugs, or a specific dollar amount with one wholesaler. Rebates are nearly impossible to trace, quantify, and assign to a particular product, and yet are an important component of the market. I am not able to correct for the measurement error caused by the existence of rebates. However, the law explicitly instructs firms to include cash rebates

in their calculation of unit prices, so a firm has no particular motive to alter its pattern of rebates.

In this paper, a "drug" refers to a specific chemical entity which may be called by its generic or brand (proprietary) name. A drug could potentially be manufactured by several firms. A firm may decide what forms and concentrations to produce; a form is solid (e.g. tablet), injectable, or other (syrup, patch, or cream). Within a form, a drug can come in different concentrations (e.g. 20 or 50 mg tablets, 1 or 2 mg/ml). Additionally, within a form and concentration choice, the firm may vary the presentation. For example, a solid can be a tablet, capsule, coated, or chewable. In the same way, an injectable can be a suspension or a solution. The final choice variable for the manufacturer is the packaging. Tablets can come in bottles of many sizes from 10 to 1,000, liquids in pre-filled syringes, vials, or cartons of vials. All these variables are included in my data; each observation is a unique combination of drug, labeler, form, concentration, presentation, and packaging. The previous literature on pharmaceuticals has largely worked with the most common (highest revenue) dosage form to avoid these complex dimensions. I hope to be able to add more depth to the analysis with the additional information.

Data

The data were collected by IMS America, a firm that provides detailed data about pharmaceutical sales in the U.S.. IMS provided the Cardiovascular subset of their *Drugstore Audit* and of their *Hospital Audit* from 1990 through 1991. These audits are created by monthly sampling of warehouse, chain pharmacy, and independent pharmacy

invoices for observations on the wholesale price. The individual prices are then combined into a weighted average that is a reported estimate of national revenue and quantity. The individual prices are never reported. The fact that the price the researcher sees is an average price is critical to the interpretation of later results. IMS takes out a large fraction of all price dispersion before the data are seen by anyone, and the data never included ex-invoice cash rebates to begin with. The Audits report revenue and quantity by presentation and manufacturer. IMS collapses all the price observations for that drug-manufacturer-presentation pair into one monthly revenue-quantity pair.

The IMS data have two important features. The first is that the IMS "manufacturer" is not always the actual manufacturer; instead it is the labeler. A manufacturer may label all or part of its own output. It may also sell all or part of its output to one or more labelers who put their own firm name on the package. The labeler, not the manufacturer, is responsible to HCFA for the Medicaid Rebate on all its products. Thus each labeler has its own AWP and rebate amount which could differ from other firms selling the identical product. I can identify the true set of manufacturers allowed to make a given drug because each one must have filed an NDA or an ANDA with the FDA. However, some manufacturers sell all their output to labelers in which case they will never show up as an IMS manufacturer. The second feature is a significant drawback of any pharmaceutical dataset: cash rebates are not included in the revenue figures; only amounts written on the invoice are tabulated in the data.

IV. Patterns in Price Dispersion

The most natural explanation for dispersion of prices across pharmaceutical products is the unequal costs of different types of packaging. The cost of putting a thousand tablets in one bottle is undoubtedly less than that of packaging 100 bottles of 10 tablets each. Some element of observed price discrimination will therefore be attributable to cost differences. I claim, however, that the amount of dispersion is in excess of what can be explained by differing costs. I predict prices in my sample with characteristics of the drug and packaging. The price that is predicted is a per milligram (or milliliter) price. The number of units sold in each observation is multiplied by the concentration of the product to arrive at a total number of units of raw material. Each drug, form, and brand has its own dummy proxying for the cost of the raw materials. The brand fixed effects capture average mark-ups over cost as well as cost variations. A *Relative Concentration* variable allows for a nonlinear relationship in price between different concentrations of the same drug. Additionally, each type of packaging gets its own dummy variable, weighted by the inverse of the number of units in the package. The price of each milligram of the drug therefore includes its portion of package costs as well as raw materials. If costs are all that make up price, the number of units in the package should turn out to be insignificant in this specification, while packaging coefficients should reflect the cost borne by each pill. I include the number of units, grouped into several categories for ease of estimation, to test the cost hypothesis.

The regression results are reported in Table II. The drug*form dummies absorb a lot of variation, contributing heavily to an R^2 of 0.829. Both the number of units in

the package and the type of packaging explain a statistically portion of variation in price.[64] The coefficient on packages of one unit implies that those packages cost far more than others: on average eighteen dollars more. Above this level, the number of pills in a bottle does not significantly affect prices of otherwise similar products. However, the coefficients on package size are nowhere near reasonable levels. Instead of being small and positive, they are large and usually insignificant. Some of the effects of the number of units in a package may be being picked up here, but some packages appear to have negative costs, while others are very expensive.

The results of the same regression on hospital prices are somewhat different. None of the number of unit variables in significant; this result is consistent with the bargaining power, information, and relatively large demand of hospitals. Packages again occasionally have negative costs. Additionally, the variation seen here is among average prices; only a fraction of the existing variation is included in the data. For the cost argument to explain all price dispersion, all the explanatory power of the dummy variables must be due to cost differences, the unexplained variation must be irrelevant, and all the unseen variation must be cost driven as well. In addition, variables that explain market structure should have no impact on the amount of dispersion, which we will see that they do in the next section.

The alternative explanation to price differences driven by costs is that firms may be discriminating among customers on the basis of elasticity of demand. The key insight

---

[64] An Analysis of Variance of the regression shows the F-tests for package and number of units reject the hypothesis that these variables do not explain price.

of Borenstein (1985) and Holmes (1989) is that the elasticity of demand for a product can be decomposed into an "industry elasticity" and a "cross-price" elasticity. The industry elasticity is based on heterogeneous consumer valuations for a product; when the price rises, consumers may decide not to buy anything at all. Cross-price elasticity, on the other hand, represents the willingness of consumers to switch between substitutable goods of different firms. Consumers respond to price changes not by dropping out of the market, but by changing products. Using Holmes' notation where ϵ^I is industry elasticity, y is industry demand, and p is price,

$$\epsilon^I(p) = -\frac{p}{y(p)} \cdot \frac{dy(p)}{dp} \tag{11}$$

Holmes analyzes a duopoly where the change in the demand for firm A's product can be decomposed into two parts,

$$\frac{\delta x_i^A(p,p)}{\delta p_A} = \frac{dy_i(p)}{dp} - \frac{\delta x_i^B(p,p)}{\delta p^A} \tag{12}$$

Converting the terms above into elasticities, he shows that the firm's own elasticity of demand breaks down into two parts,

$$\epsilon_i^F(p) = \epsilon_i^I(p) + \epsilon_i^C(p) \tag{13}$$

(Here the superscript F stands for total Firm elasticity and C for Cross-price elasticity.)

Since monopolies have no perfect substitutes in this industry, they will employ industry type price discrimination relatively more than generics and brands sharing a market.[65] The cross-price elasticity between a brand and its generic substitute varies by consumer, so we expect firm elasticities of demand to vary more widely for competitive brands than patented brands. That in turn implies that competitive brands and generics will price discriminate more than other firms if they are trying to serve their whole potential market.

What empirical measures of dispersion are available? I cannot measure a large segment of price dispersion with the information included in the IMS data; the dispersion due to prices differing across customers within a sector is hidden. However, hospital and drugstore data are listed separately, so customers are divided into two main groups. Additionally, some of the dispersion which is affected by the rebate law comes from differences in price per pill due to different package sizes. The IMS data does include this type of price dispersion. If a drug-presentation-labeler comes in several different sized packages, each size is listed separately. The per pill (or per ml) price will likely be different across package size within a product and labeler. These price inequalities provide the variation that allows me to form a measure of price discrimination. Of course, the measure captures only one type of price discrimination occurring in the market, package-size discrimination. If firms specialize in different types of discrimination, only those firms that use the package-size type will show up as

[65] As mentioned above, a potential problem for my analysis is that cross-price elasticities between the monopoly product and alternative therapies certainly exist. Future research will include gathering a good quantitative measure of therapeutic substitutes.

discriminating. However, that advantage of observing package-based price discrimination is that it is affected by the rebate law and so its movements can be studied over time.

To quantify the amount of price dispersion within an observation, I use two weighted coefficients of variation, CV and CVd. The first is constructed using the prices described above with quantity weights; it is simply the standard deviation of prices over the mean price. This CV is constant within a drug-form-labeler-concentration and is the same for equivalent observations in both the hospital and drugstore sectors. The variations in price (within the drug-form-labeler-concentration) that are used to created CV come from multiple packaging quantities and observations in both the drugstore and hospital audits. The second coefficient of variation, CVd, is created using only the variation from multiple packaging quantities within the *drugstore* audit. It measures price dispersion within the drugstore market only. Both CV variables were calculated for each month in the years 1990 and 1991. The values of the variable CV are bunched up close to zero with only a few observations having large CVs. A substantial proportion of observations (41% percent for CV, 36% percent for CVd) have CV equal to zero.[66]

We would like to determine whether the value of CV or CVd is affected by the Medicaid legislation. To give a rough idea of movements in the data, I calculate summary statistics for the two coefficients of variation. Competitive brands discriminate the most, followed by ANDA holders. Off-patent brands have a much lower amount of

[66] CV equal to zero occurs when a presentation-labeler-concentration appears in the dataset only once and so only has one price.

price dispersion than all the other groups.[67] Table IV contains the summary statistics by class, before and after the legislation change, for both CV and CVd. Although the standard deviations make changes in the means statistically insignificant, CVd declines after the MFN legislation is passed in all classes of drugs except generic distributors. In particular, measured price dispersion falls considerably for drugs in the competitive brand class and ANDA holders. The competitive brand class has the highest level of price dispersion, with generics next. Drugs with no exact substitutes, off-patent brands and patented brands, show the least amount of price dispersion both before and after the legislation change.

One difficulty in trying to relate price dispersion to competitive characteristics is that each product has its own level of consumer heterogeneity. It may be the case, for example, that one drug treats several different conditions (in fact, most do) and has different substitutes for the different conditions. Thus the distribution of elasticities of demand for each presentation of each drug could be different. (This feature of the data could explain the large standard deviation in CV noted above.) I therefore allow each presentation in the dataset to have a separate intercept, or initial level of dispersion in the following regressions.

I regress CV and CVd on a time trend, the presentation intercepts, and a variety of shift terms. The first specification includes one shift for all brands and another for all generics. The next specification allows each class of drug to have its own time trend

---

[67] One could perform a regression of CV on a variety of explanatory variables within one month, rather than across months, to help determine the causes of dispersion.

and own shift. All the shifts take effect in January 1991. If the shifts' signs are significant, the legislation made an impact on the price dispersion observed in the IMS data.

Table V shows that the amount of price dispersion across drugstore and hospital markets actually rises significantly for branded products in 1991. However, with the drugstore market dispersion falls although the coefficient is not significant at conventional levels. When brands and generics have separate time trends in the third and fourth columns, the generic upward shift and the brand fall become more significant. In the first columns of the next table, Table VI, each class is given its own time trend and shift. Only the competitive brand class experiences a decline in drugstore price dispersion; in contrast its cross-market CV rises significantly. This result is consistent with the means observed in Table IV, and accords with the idea that competitive brands are trying to sell to a group of consumers with widely differing cross-price elasticities. The other significant shifts in dispersion occur in the drugstore market. ANDA holders reduce dispersion and generic distributors increase it considerably. The question of why generic distributors might increase price dispersion after legislation that does not directly affect them is a mystery.

Other characteristics besides the four competitive groups included above may explain the change in price dispersion from 1990 to 1991. I include explanatory variables that measure the structure of a firm's market in the second half of Table VI. The information about the different markets is available due to thorough regulation of pharmaceutical production by the FDA. It is straightforward to identify which firms are

participating, or have permission to participate, in each market by examining the FDA's Approved Drug Products and Therapeutic Equivalents.[68] This publication details which ANDA's have been granted; it reports the exact concentration, form, date of approval, and firm receiving the ANDA. Table III gives descriptive statistics for the variables discussed below.

I created the variable *Number of ANDAs* for each observation; it is the total number of ANDA's issued for that strength and form of that product.[69] *ANDA* is a dummy variable taking the value one when the observation's manufacturer holds an ANDA. *Duopoly* is assigned a one if the drug in question was marketed as a brand by two separate firms.[70] I create a variable *Hospital Share* equal to the fraction of hospital and drugstore revenue a drug-form earns from hospitals. I use a spline function to allow for nonlinear effects of hospital share. *Hospital1* equals *Hospital Share* when *Hospital Share* itself is between 0 and 0.3. Similarly, the range for *Hospital2* is between 0.3 and 0.7, and that of *Hospital3* is between 0.7 and 1. If the hospital and drugstore markets function very differently, the proportion of output sold in one or the other will affect the extent of price dispersion. The hospital share variables are used only in the CV specification.

---

[68] It is useful to think of the firm as holding a portfolio of ANDA's or NDA's, any of which it may exercise at a given time. This analogy is possible because capital equipment is form specific; it can usually make any drug within a form.

[69] A generic firm must submit a separate ANDA for each concentration of each form of a drug it would like to manufacture.

[70] The two firms might have both licensed the drug, or they might have discovered it simultaneously.

Table VI reports the regression with the additional variables interacted with the shift from 1990 to 1991. The shift associated with competitive brands is negative and significant for both CV and CVd now. The market structure variables explain dispersion well. An increase in the *Number of ANDAs* raises price dispersion. A higher hospital share also raises price dispersion across sectors after the MFN. Although hospital prices are exempt from the law, if drugstore prices change relative to hospital prices, dispersion could alter in response to the policy. Here, an increase in competition and "perfect" substitutes increases price dispersion. Duopolies shifted to less drugstore price discrimination after the MFN passed; the existence of a close substitute may have caused them to discriminate more initially - like the competitive brands.

A firm which is a manufacturer and distributor with its own ANDA sharply reduced package-type dispersion in both sectors after the legislation. We know that generics were discriminating on the basis of package size more than patented and off-patent brands before the MFN rules were passed. The response here is not due to the best price provision, as it does not apply. However, the average price provision could affect generics, or the change could be a response to brands' altered behavior. Only generics appear to have higher dispersion in the drugstore sector after the legislation is passed. The brands with no competitors do not change their levels of measured price dispersion. These findings confirm the within and between firm effects discussed above, and suggest that the prices and quantities of drugs in the competitive brand and ANDA classes will be most affected by the legislation.

V. Estimation of Price and Quantity Shifts

How did prices and quantities change as the market responded to Medicaid "best price" legislation? Basically, a firm which is price discriminating will no longer want to sell any output at its lowest prices, because that would trigger the MFN. The strategic effect of an MFN clause is to encourage "puppy dog" behavior in an oligopolistic market.[71] The brand's competitors (therapeutic substitutes or generics) will react to the strategic effect by raising price. The different treatment of the hospital and drugstore markets means that the effects of the law will not be felt equally in the two markets.

I examine the behavior of price and quantity over the two years surrounding January 1, 1991, the first effective date of the new policy. IMS is careful to record wholesale transfers in the month in which they actually occur, so error from data lags is unlikely. However, there are reasons to suspect the shift might not occur cleanly on January 1, 1991. The rebate program was devised and passed so quickly that many state Medicaid offices were not organized to collect the rebates during the first quarter of 1991. Firms were not obligated to pay the rebates if the states did not have the data to show how much was owed.[72] Compliance among firms improved over the first quarter of 1991 as remaining states implemented the rebate scheme. In addition, firms themselves might not have adjusted instantaneously on January 1, 1991. Those with better management or analysis of the strategic consequences of their prices might use a

---

[71] The taxonomy in Tirole (1988) p327 describes how a firm would like to underinvest if investment makes it "tough" when prices are strategic complements and the firms are competing in price. Such a strategy is called "puppy dog."

[72] However, the larger firms estimated the amounts owed and paid anyway.

new pricing strategy more promptly than other firms; we could observe a short time lag in behavior adjustment. I report the results from a model of a quarter lag in price adjustment. The model which best fits quantity movements is more complex and is discussed below.

The original dataset contains every cardiovascular item tracked by IMS. I drop those observations for which I cannot to identify the competitive class and also those with revenue or quantity values below two. Since IMS rounds to the nearest integer, a price can be significantly mismeasured if the numerator or denominator used to construct it is recorded as a one. I also modify all the datasets further by dropping presentations without more than twelve months of data; percent change estimates can become unreasonable if only a few data points are available. Many revenue or quantity observations are missing or zero for a number of months. I create two smaller datasets which are variations of the full dataset. In the first, I drop all observations missing revenue or quantity. Finally, I make the dataset balanced by removing any observation that does not have *both* hospital and drugstore prices. This step allows better comparisons between hospital and drugstore movements. The second dataset is the complement of the first; observations that are not included in the balanced dataset because of missing information comprise the unbalanced dataset.

The basic regression is OLS using log price or log quantity as the dependent variable. Taking the log of price or quantity allows the shift and trend terms to be expressed in percent changes and therefore be applicable across presentations and drugs. The explanatory variables include presentation dummies; each presentation of each drug

has its own intercept. Each drug-class has its own time trend which is forced to be equal across presentations of that drug-class. The critical variables are five class dummies which take the value one if the drug is a patented brand, an off-patent brand with no generic competition, a competitive brand, an ANDA holder, or a generic distributor, respectively. Each observation falls into only one class. The estimated coefficients of the class shifts (β_2 in the equations below) will capture the change in price or quantity that is common to all observations in that class.

$$\log P_{idc} = \beta_0 Intercept_{idc} + \beta_1 Trend_{dc} + \beta_2 MFNLaw_c + \epsilon_{idc} \qquad (14)$$

and,

$$\log Q_{idc} = \beta_0 Intercept_{idc} + \beta_1 Trend_{dc} + \beta_2 MFNLaw_c + \epsilon_{idc} \qquad (15)$$

Where i is presentation, d is drug, and c is class. I analyze hospital and drugstore markets separately; P is either drugstore price or hospital price, for example. The results from analysis of the two markets are reported in different columns of the tables below.

Prices and quantities are not observed with equal accuracy across observations. We suspect that the variance of both reported revenue and reported units is inversely related to the level of revenue or quantity. Presentations that have a small level of sales, but are included in the dataset will have much more measurement error than presentations

that earn large amounts of revenue. Thus, OLS standard errors will not be correct. Instead, I use Weighted Least Squares (WLS) and weight by the share of the presentation's revenue in its drug-class revenue.

The package size of a product could also influence the price response to the law. Large packages usually have a lower unit price. If package size is used as a way to discriminate among customers, the largest packages will be sold at the lowest prices and will draw the "best price" provision. In such a case, firms will want to raise the price of the largest package sizes, but not others. Under this hypothesis, smaller packages with higher unit prices will not see as large a rise. *Maximum Size* is defined to be a one if the observation's package size is the largest available for that drug in the dataset. Alternatively, the largest buyer might get the same low price across whatever package sizes it wants, in which case the variable *Maximum Size* will not be significant. The equation for log price which includes *Maximum Size* is,

$$\log P_{idc} = \beta_0 Intercept_{idc} + \beta_1 Trend_{dc} + \beta_2 MFNLaw_c + \beta_4 MaximumSize_c + \epsilon_{idc} \quad (16)$$

The price specifications in Table VII and Table VIII include a simple January shift interacted with each class. The time trend omits the period January to March 1991 for each class of drug to allow for market adjustment. The tables report the specifications above with the addition of a *Maximum Size* variable interacted with each

class using both the balanced and unbalanced datasets. In Table VIII two summed coefficients are reported for each class. The upper coefficient is the total price shift term for a package of smaller size. The lower number is the price shift for a package of *Maximum Size*. The results here demonstrate the effect of the MFN on prices. Prices for patented brands and competitive brands rise more in the drugstore sector than in the hospital sector. The third column reports the results of the differences in differences experiment. The differences which are strongly significant are competitive brands, as expected, and generics, where prices actually rise relatively for hospitals and fall for drugstores. The difference between hospitals and drugstores is not significant at conventional levels for patented brands.

The estimate of the *Maximum Size* coefficient on its own is insignificant for patented and competitive brands. This result indicates that brands are changing prices in a similar fashion across packages. The *Maximum Size* variable has the most effect on large generic packages which decline in price more than smaller sizes. The coefficient on the *Maximum Size* shift becomes negative for all competitive classes when the balanced dataset is used; the results in Table VIII are not as consistent with theory as those from the unbalanced dataset.

The price rises in the hospital sector in both tables are a puzzle as they are exempt from the legislation. One simple explanation is that some firms' quarterly price rises occur in January and some of that is being picked up in this specification.[73]

[73] The January price rise explanation does not work well in the next table. It uses a January shift with a continuous time trend and the shift coefficients are not nearly as large.

Alternatively, firms might be worried about arbitrage across sectors and increase list prices to hospitals while increasing rebates as well. The balanced dataset described above only includes products that are sold both to hospitals and drugstores during this period. If arbitrage or a list pricing-rebate tradeoff is operating, the differences between estimated price shifts in hospitals and drugstores using the balanced dataset should decline compared to those in Table VII, which is what we observe.[74]

Table VIII reports a regression using the balanced data where the shift terms and the *Maximum Size* terms are reported separately. The bottom portion of the table sums all relevant coefficients for an observation of *Maximum Size*. This number is equivalent to the summed Maximum Size coefficients in Table VII. We see that the absolute price rises in the different categories are similar, but the difference between hospital and drugstore has reversed itself; hospital price rises are larger. The comparison between drugstore and hospital shifts depends on whether the sample of drugs contains products that have constrained prices because they are sold in the other market.

The pattern of estimated shifts from the balanced dataset does not match the theory as closely as that of the unbalanced data. If we believe that firms profit-maximize always, these results are discouraging. However, it might be the case that firms differ in their information about the law and its strategic impact; some firms might spend more

---

[74] A more complex alternative story might be that the quasi-rents pharmaceutical firms earn are limited by political pressure. Suppose the government required a transfer of rents from the industry to the government to help reduce expenditure. The industry would then no longer be keeping as much profit and could raise prices to restore quasi-rents to their original levels without provoking public outcry. The policy introduction would then provide a focal moment for pharmaceutical firms to raise prices in markets otherwise unaffected by the MFN policy. Of course, political scrutiny might have had the opposite effect of reducing price increases, as has been the case recently.

resources on determining their optimal pricing strategy. These explanations would suggest that firms' price response to the rebate law varies across firms.[75]

To get a rough idea of whether firms differ, I isolate larger firms and check to see if they behaved differently than small firms. The variable *BigFirm* is a dummy which takes the value one if the recorded labeller or its parent company is listed in the Fortune Global 500 or Fortune US 500 in 1990. The idea is that the marginal benefit of figuring out the optimal response to the MFN legislation is large for *BigFirms* compared to others. Therefore, they might respond more optimally than smaller firms. I use the same price and quantity specifications as above, but I restrict the observations to those having BigFirm equal to one.

Table IX reports the results of the *BigFirm* regressions. The competitive brand shifts are larger than in the balanced data results; these firms are raising price to some consumers. The maximum size summed coefficients for patented and competitive brands are more similar to the unbalanced data results, although again, hospital and drugstore coefficients are often roughly equal. Overall, the results for *Big Firms* are tighter and support the theory better.

Quantity behavior is slightly more difficult to model. Inventory effects, both buying early and buying late, as well as different effects for different package sizes will affect the time path of units purchased. If we believe that agents are rational and expect the price rises that the law will bring, quantity changes will not follow the same time

[75] Also, if firms differ in their initial pricing scheme, they will have different optimal responses to the law, even if they all have exactly the same information.

pattern as price changes. Consumers that realize prices will rise on January first will purchase additional quantities before that date. Manufacturers renegotiated low-price contracts that would have carried over into the new regime; those customers who found their contract cancelled would certainly have had the information and time to arrange purchases optimally.[76] In particular, we might expect customers to purchase in December 1990 rather than January 1991.

The first specification reported in Table X uses the full dataset and a specification similar to the one used for price above. The difference is that the shift term occurs in December 1990, and the time trend omits that month also. The result that comes through clearly is that competitive brands lost a large percentage of sales in the drugstore market, between twenty and forty-five, while despite similar-sized price rises, patented brands did not lose drugstore sales. The readily-available generic substitute for a competitive brand likely accounts for the big quantity loss compared to patented brands. In particular, the estimates indicate that the largest size package lost about forty-five percent of its volume during this quarter. Large sizes of ANDA and generics gained significant drugstore sales after the MFN legislation passed, as we would expect when brand prices rise. Generics, though, are not being bought early by consumers, so we should not expect those two groups to fit the December specification particularly well.[77]

In Table XI I estimated the same specification using the balanced dataset to see how the results are affected. Quantities sold to hospitals are accurately reported so the

---

[76] However, the customers with contracts were HMOs and hospitals, groups that are either not in my dataset or unaffected by the law.

[77] Later in the paper generics are analyzed in a separate specification.

drugstore-hospital comparison should be valid. The results for Maximum Size packages at the bottom of Table XI confirm the theory. Almost no quantity movement occurred in the hospital market, while competitive brands lost sales in drugstores and generics gained them. The results from the same specification run on the *BigFirm* sample follows in Table XII. Again, the results are a little bit tighter, especially the ANDA results, although patented brands are unexpectedly increasing price in the hospital market.

The empirical results thus far have focussed on firms that have both rivals (therapeutic substitutes) and price discriminate. In order to isolate the strategic effect I examine a set of firms that are not themselves subject to the MFN, but compete in an oligopoly with a firm that is subject to the MFN. Table XIII focusses on the strategic effect by reporting two regressions for generic prices and quantities only. One group contains markets with three or fewer generic firms.[78] The market is small enough for the brand's constraint by the MFN to have an effect on generic behavior. In contrast, markets with many generic firms ought to be practicing vigorous Bertrand competition. The fact that the brand faces an MFN is insignificant when competition is provided by many other players. Notice that both sets of generic firms are subject to the average price provision of the Rebate Law on products they sell to drugstores.

The only other factor that might affect the results of the experiment is capacity constraints. If there are only two firms which have ANDAs for a product and demand increases sharply, the two firms might have trouble meeting demand. Thus a comparison

---

[78] If markets with only one generic form a group, it contains only four drugs and the results could easily be the product of idiosyncratic factors. In contrast, the group which selects markets with three or fewer ANDAs includes 21 different drugs.

of the two drugstore-hospital differences will yield an estimate of the strategic plus capacity constraint effects.

Table XIII shows the difference between markets that have three or fewer approved generic manufacturers and those that have four or more. Both ANDA holders and generic distributors are included; the ANDA estimates may be more reliable since generic distributors are heterogeneous and there may be agency problems that are unobserved to the econometrician. ANDA holders in few-ANDA markets do not change price significantly with the legislation. However, ANDA holders in large markets shift drugstore price down by 40%. The hospital market displays a similar pattern; ANDA prices drop only very slightly in the small markets, but drop by 10% in the large markets. The generic results in the top half of the table are erratic. Prices seem to be declining in small drugstore markets, rising in small hospital markets and are otherwise unchanged.

For the quantity specification, I use a shift that lasts from February to March 1991; consumers may be able to buy in December to cover their January demand for brands, but by February they should be buying generics. The markets with three or fewer generic manufactures show large positive quantity shifts for both ANDAs and generics; this result is consistent with the quantity loss by competitive brands. However, the generics appear to be gaining some quantity in the hospital market also, which contradicts the theory. Quantities sold by ANDA or generic firms in markets with many ANDAs show smaller shifts than those above, and maximum size shifts of approximately zero.

Table XIV shows the same specification using the full dataset. The results are somewhat more reasonable, but the conclusion is unchanged. Price behavior of generics in small markets does not indicate they are responding to the strategic effect or the capacity constraints effect. Quantity sold by generics increases in the drugstore market more than in the hospital market and firms in small markets have a larger increase than those in large markets.

In conclusion, ANDA holders are increasing quantities in those markets where there are relatively few suppliers; quantity shifts are larger in drugstore markets than in hospital markets. The difference in quantity shift between the drugstore and hospital sectors gives support to the notion that hospital prices are not rising as much as the estimates here indicate. Evidence for a strategic response by generics to the brand's price increases is weak. Prices shift up relatively more in the small drugstore market. However, the coefficients are imprecisely estimated, and the effect does not hold for generic distributors.

VI. Conclusions

The evidence in this paper suggests that not all price dispersion in the wholesale pharmaceutical market is cost driven. Competitive conditions and the nature of substitutes results in different levels of price discrimination for different pharmaceutical products. Brands undertaking a great deal of price discrimination are most affected by the MFN law. The results show that competitive brands lose quantity quite dramatically and raise average price after the law takes effect. The average price of brands facing

generic competition increased by four to nine percent and those manufacturers lost about forty percent of quantity sold to drugstores.

The evidence that generic manufacturers in a market with three or fewer generic manufacturers raised average price after the MFN law took effect is weak. Average ANDA drugstore prices increased compared to hospital prices, while drugstore quantity increased dramatically. Although this group had to rebate 1.5 percent (10 percent of 15 percent) of their sales to Medicaid, the quantity increase combined with non-decreasing prices offset the rebate payments in the drugstore market. It is not too surprising that the strategic effect is difficult to estimate; the MFN experiment is weak because only 15 percent of a firm's product was subject to the MFN, the MFN ceased to bind after a firm's lowest price was above 88 percent of its average price, sales to HMOs are not included in the data, and measurement error may be significant.

Additionally, my dataset does not have enough examples of single generic firms in a market to estimate the price increase which may have occurred there. However, theory would predict that a sole generic would raise price more than a generic facing competition. Thus, firms manufacturing the only generic in a market may have gained even more from the new law. The profit gain could, of course, be temporary as entrants would arrive about two years after the market became attractive. Alternatively, generics in small markets might be practicing limit pricing; their prices do not rise precisely because they do not want to attract entry. In addition, the generic quantity and price response (such as it is) disappears when the number of generics in the market is large, indicating that the effect of MFN dissipates with competition.

I also find evidence that the best price provisions had an effect on patented brands. If the brands were price discriminating only a little when the MFN law came into effect, as measured CV suggests here, then the within-firm effect must have been weak. Any price rise must therefore have been due to strategic interaction between firms. Price competition softened and average prices increased by three to five percent in the drugstore market. Alternatively, if patented brands were discriminating a lot, the price increases could have been caused by a within-firm effect as well as the strategic effect. However, if these two effects were driving the price changes, hospital prices should not display an absolute rise as they do. Firms changing list prices on invoices while compensating with hospital rebates may be the explanation for the hospital rise. Producers of patented brands may well have gained from the imposition of the MFN clause; the result for each manufacturer would depend on its own price distribution of sales, its distribution of Medicaid sales, and the extent of quantity loss due to higher prices.

Although the Federal and state governments saved several hundred million dollars in pharmaceutical expenditure, total national expenditure on pharmaceutical products did not fall by that amount. Rather, increased prices due to the strategic effects of the MFN law partially compensated firms for Medicaid rebates. Products such as patented brands with few good substitutes and ANDA manufacturers in small markets may even have benefitted in total from the law. The legislators' objective when writing OBRA 90 may simply have been to reduce government expenditure on Medicaid. However, it is important to note that, in theory, this objective could have backfired. If the strategic

effect had been strong enough, it could have actually outweighed the rebate given to Medicaid. In that case the government's purpose would have been defeated. In fact, as we might guess, the strategic effect turned out not to be sufficiently strong to overcome the direct effect, but certainly offset it.

If legislators wished to take into account the general welfare of the nation's consumers, then the MFN clause is still an inappropriate means of reducing Medicaid expenditure. The MFN caused prices to rise, which caused consumption to drop or shift to other products. For example, if consumers were previously purchasing a competitive brand at a low price rather than a generic, they received more consumer surplus from the brand. The MFN clause effectively removed the competitive brand option. The welfare loss to consumers is composed of a loss of consumer surplus due quantity reductions in patented brands and generics plus the surplus loss due to switching products. In a market economy the government should ideally intervene in order to promote competition and protect the welfare of consumers. It is somewhat ironic that in this case government regulation produced less competition rather than more and reduced consumer welfare.

Bibliography

Bai, Jushan (1993) "Estimation of Structural Change Based on Wald-Type Statistics", MIT mimeo.

Borenstein, S. and Rose, N.L., (1992) "Competition and Price Dispersion in the U.S. Airline Industry," forthcoming *Journal of Political Economy*, August 1994

Borenstein, S., (1985) "Price Discrimination in Free-entry Markets," *Rand Journal* 16:3:380-97.

Caves, R., Hurwitz, M. and Whinston, M., (1991) Patent Expiration, Entry, and Competition in the U.S. Pharmaceutical Industry," *Brookings Papers on Economic Activity*, Special Issue:1-48.

Cooper, Thomas E. (1986) "Most-favored-customer pricing and tacit collusion," *RAND Journal of Economics* 17:3:377-388.

----, (1991) "Most-favored-nation pricing policy and negotiated prices," *International Journal of Industrial Organization*:9:2:209-223.

Holmes, Thomas, (1989), "The Effects of Third-Degree Price Discrimination in Oligopoly," *American Economic Review* 79:244-250.

Holt, Charles and Scheffman, David (1987) "Facilitating practices: the effects of advance notice and best-price policies," *RAND Journal of Economics*:18:2:187-197.

Katz, M. (1984) "Price Discrimination and Monopolistic Competition," *Econometrica* 52:6:1453-71.

Omori and Yarrow (Spring 1983) "Product Diversification, Entry Prevention, and Limit Pricing," *Bell Journal*:13:1:242-48

Png, I.P.L. (1991) "Most-Favored-Customer Protection versus Price Discrimination over Time," *Journal of Political Economy*:99:9:1010-1028.

Png, I.P.L. and D. Hirshleifer (1987) "Price Discrimination through Offers to Match Price," *Journal of Business*:60:365-383.

Salop, S. (1977) "The Noisy Monopolist: Imperfect Information, Price Dispersion, and Price Discrimination," *Review of Economic Studies*:44:3:393-406.

---, (1986) "Practices that (Credibly) Facilitate Oligopoly Coordination" in J. Stiglitz and

F. Mathewson eds., New Developments in the Analysis of Market Structure. MIT Press, Cambridge.

Salop, S. and Stiglitz, J., (1982) "The Theory of Sales: A Simple Model of Equilibrium Price Dispersion with Identical Agents" *American Economic Review* 72:5:1121-1130.

Tirole, Jean, (1988) The Theory of Industrial Organization, MIT Press, Cambridge.

Varian, Hal. (1985) "Price Discrimination and Social Welfare" *American Economic Review* 75:870-75.

Table II: Cost Regression			
Dependent Variable:		Drugstore price per milligram (or ml) in January 1990	Hospital price per milligram (or ml) in January 1990
Number of Units in a Package	1	18.13 (4.41)	3.352 (2.70)
	2	2.551 (2.55)	1.801 (1.96)
	3-6	-0.1156 (1.83)	1.065 (1.66)
	7-10	-0.4462 (1.68)	1.016 (1.56)
	11-30	0.1736 (1.41)	1.538 (1.47)
	31-99	0.6640 (1.38)	1.737 (1.46)
	100	0.1047 (1.30)	0.2165 (1.37)
	101-500	-0.0920 (1.31)	-0.0423 (1.36)
	501-1000	3.467 (2.26)	0.6967 (2.80)
	1001+	-0.0476 (1.30)	-0.0175 (1.36)
Package	unit dose pack	33.54 (50.67)	-22.67 (43.0)
	unit dose bottle	-43.74 (6.33)	-33.62 (6.75)
	vial	-13.07 (4.25)	-1.307 (2.12)
	vial + IV	2.990 (2.29)	2.092 (1.81)
	ampule	-0.9775 (6.51)	-0.9796 (5.09)
	syringe	5.795 (4.65)	-2.625 (6.93)
	unit dose syringe	3.424 (84.3)	13.41 (269)
	bottle with dropper	-81.77 (47.7)	-34.53 (48.8)
	box	56.03 (4.84)	1.211 (27.9)
	tube	-1.486 (4.52)	14.98 (2.87)
	bottle	-8.189 (15.4)	-19.46 (18.9)
Concentration		-0.0054 (.0025)	-0.0040 (.0022)
		Adj R^2 = 0.829 N = 2198	Adj R^2 = 0.831 N =1410

Table III. Summary Statistics
Each Drug Weighted Equally

Balanced Dataset					
	N	Mean	Min	Max	Std Dev
Relative Size of Package	28225	0.547	0.005	1	0.428
Number of ANDAs	28225	5.12	0	23	5.91
-If Number of ANDAs>0	23128	8.09	1	23	5.58
Annual Revenue of Drug-Form	28165	6508549	186	3.17E+08	2.37E+07
Duopoly	28225	0.100	0	1	0.300
Share sold to Hospitals	28225	0.171	0.006	0.998	0.321

Dataset:	Balanced	Full	Unbalanced
Total Number of Observations	28225	80112	51887
Total Number of Presentations	1352	3338	2649
Number of Obs that are Patented Brand	4074	5160	1086
Number of Obs that are Off-Patent Brand	1023	1440	416
Number of Obs that are Competitive Brand	5902	7752	1850
Number of Obs that are Generic Distributors	11440	43824	33872
Number of Obs where labeler has ANDA	6115	16752	10318
Number of BigFirm Observations	7669	14736	5972
Total Number of Drugs	87	112	87

Table IV: Coefficient of Variation
Balanced Dataset

Whole Sample 1990-91	N	Mean	Min	Max	Std Dev
CVd: patented brands	4074	.001052	0	.257958	.00665
CV: patented brands	4074	.001815	0	.23379	.00841
CVd: competitive brands	5902	.02858	0	10.051	.3386
CV: competitive brands	5902	.02560	0	2.3910	.09960
CVd: off-patent brands	1023	.000229	0	.15075	.00364
CV: off-patent brands	1023	.000285	0	.19333	.00440
CVd: ANDAs	6115	.01408	0	2.4211	.09161
CV: ANDAs	6115	.01478	0	2.3910	.07652
CVd: generic distributors	11440	.00600	0	1.2134	.04966
CV: generics distributors	11440	.00697	0	.86130	.04984

Means Before and After MFN[79]	CV		CVd	
	1990	1991	1990	1991
Patented Brands	.00204 (.01065)	.00161 (.00563)	.00106 (.00738)	.00104 (.00590)
Competitive Brands	.02812 (.10349)	.02296 (.09528)	.03446 (.38380)	.02243 (.28356)
Off-Patent Brands	.00024 (.00116)	.00032 (.00608)	.00025 (.00479)	.00021 (.00194)
ANDAs	.01761 (.08064)	.01207 (.07224)	.02001 (.11460)	.00838 (.06137)
Generic Distributors	.00739 (.05360)	.00656 (.04594)	.00418 (.04106)	.00774 (.05664)

[79] Standard deviations are in parentheses.

Table V: Changes in Measured Price Dispersion
Balanced Dataset[80]

Dependent Variable:	CV	CVd	CV	CVd
	N=28165	N=28165	N=28165	N=28165
Time	-.000057 (.000028)	.000028 (.00003)	---	---
Time - Brand	---	---	-.00013 (.00003)	.00007 (.00004)
Time - Generic	---	---	.00011 (.00005)	-.00007 (.00005)
Shift - Brand	.00224 (.00043)	-.00050 (.00046)	.00269 (.00045)	-.00075 (.00048)
Shift - Generic	.00078 (.00064)	.00016 (.00068)	.00010 (.00066)	.00199 (.00071)
Constant	.00760 (.00030)	.00217 (.00032)	.00760 (.00030)	.00217 (.00032)
Presentation Intercepts?	yes	yes	yes	yes
Adj R-squared	0.703	0.353	0.704	0.353

[80] Standard errors are in parentheses. Log(CV) might be a better specification, but I do not use it due to the large number of observations with CV=0.

Table VI: Changes in Measured Price Dispersion Balanced Dataset				
Dependent Variable:	CV	CVd	CV	CVd
Shift - Duopoly	---	---	.00738 (.00144)	-.00451 (.00148)
Shift - Number of ANDAs	---	---	.00171 (.00012)	0.00023 (.00012)
Shift - Share Hospital0	---	---	-.01612 (.03106)	---
Shift - Share Hospital1	---	---	.00132 (.00368)	---
Shift - Share Hospital2	---	---	.00748 (.00113)	---
Time - Patented Brand	9.27E-06 (.00008)	3.95E-06 (.00008)	9.00E-06 (.00008)	3.95E-06 (.00008)
Time - Competitive Brand	-.00038 (.00006)	.00019 (.00006)	-.00039 (.00006)	.00019 (.00006)
Time - Off-Patent Brand	1.17E-06 (.00005)	4.5E-06 (.00005)	9.01E-08 (.00005)	4.50E-06 (.00005)
Time - ANDA	.00003 (.00006)	-.00003 (.00007)	.00009 (.00006)	-.00003 (.00007)
Time - Generic Distributor	.00003 (.00008)	-.00010 (.00009)	-.00001 (.00008)	-.00010 (.00009)
Shift - Patented Brand	-.00022 (.00107)	.00043 (.00114)	-.00161 (.00130)	.00048 (.00114)
Shift - Competitive Brand	.00798 (.00078)	-.00181 (.00083)	-.00321 (.00106)	-.00219 (.00098)
Shift - Off-Patent Brand	9.26E-06 (.00065)	-1.46E-06 (.00069)	-.00128 (.00085)	-1.46E-06 (.00069)
Shift - ANDA	.00030 (.00087)	-.00429 (.00093)	-.01346 (.00130)	-.00526 (.00124)
Shift - Generic Distributor	-.00019 (.00104)	.01072 (.00111)	-.01279 (.00135)	.00987 (.00131)
Constant	.00779 (.00030)	.00218 (.00032)	.00776 (.00030)	.00218 (.00032)
Presentation Intercepts?	yes	yes	yes	yes
Adj R-squared	0.704	0.356	0.709	0.356

(Standard errors are in parentheses. The number of observations is 28165.)

Table VII: Price Shifts from January to March 1991
Unbalanced Dataset
Weighted by Revenue and Inverse of Number of Observations per Drug-Class[81]

Summed Coefficients Reported for Two Package Size and Two Sectors

	Package Size	Drugstore	Hospital	Difference
Patented Brand	Smaller Size	0.0724 (.0151)	0.0393 (.0153)	0.0332 (.0214)
	Maximum Size	0.0767 (.0160)	0.0516 (.0177)	0.0251 (.0239)
Competitive Brand	Smaller Size	0.0938 (.0051)	-.0005 (.0083)	0.0943 (.0098)
	Maximum Size	0.0925 (.0066)	0.0151 (.0112)	0.0775 (.0130)
Off-Patent Brand	Smaller Size	0.0618 (.0086)	0.0523 (.0184)	0.0095 (.0203)
	Maximum Size	0.0269 (.0106)	0.0131 (.0217)	0.0138 (.0242)
Generic	Smaller Size	-3.0E-5 (.00081)	0.0905 (.0202)	-0.0906 (.0219)
	Maximum Size	-0.0620 (.0067)	-.0272 (.0177)	-0.0348 (.0189)

[81] Standard errors are in parentheses.

Table VIII: Price Shifts from January to March 1991
Balanced Dataset
Weighted by Revenue and Inverse of Number of Observations per Drug

		Drugstore	Hospital
Patented Brand	Common Shift	0.0570 (.0133)	0.0691 (.0223)
	Maximum Size Shift	-.0474 (.0100)	-.0108 (.0167)
Competitive Brand	Common Shift	0.0526 (.0117)	0.0187 (.0197)
	Maximum Size Shift	-.0120 (.0054)	-.0536 (.0091)
Off-Patent Brand	Common Shift	0.0804 (.0063)	0.0775 (.0106)
	Maximum Size Shift	-.0426 (.0058)	-.0678 (.0097)
ANDA	Common Shift	-.1543 (.0145)	0.0412 (.0244)
	Maximum Size Shift	-.0840 (.0066)	0.0879 (.0110)
Generic	Common Shift	-.0317 (.0170)	0.0998 (.0285)
	Maximum Size Shift	.0847 (.0075)	0.0879 (.0126)
Maximum Size - Summed Coefficients	Patented Brand	0.0096 (.0112)	0.0582 (.0167)
	Competitive Brand	0.0405 (.0116)	0.1563 (.0091)
	Off-Patent Brand	0.0378 (.0050)	0.0240 (.0097)
	ANDA holder	-0.2383 (.0151)	-.0267 (.0110)
	Generic Distributor	0.0530 (.0175)	0.1877 (.0126)

Table IX: Price Shifts from January to March 1991
Big Firms Only
Balanced Dataset
Weighted by Revenue and Inverse of Number of Observations per Drug

		Drugstore	Hospital
Patented Brand	Common Shift	0.0593 (.0181)	0.0653 (.0220)
	Maximum Size Shift	-.0098 (.0170)	-.0233 (.0216)
Competitive Brand	Common Shift	0.0793 (.0174)	0.0818 (.0218)
	Maximum Size Shift	-.0473 (.0103)	-.0029 (.0131)
Off-Patent Brand	Common Shift	0.0691 (.0111)	0.0829 (.0133)
	Maximum Size Shift	-.0274 (.0104)	-.0451 (.0132)
ANDA	Common Shift	-.0459 (.0205)	-.0214 (.0261)
	Maximum Size Shift	0.0327 (.0185)	-.0819 (.0235)
Generic	Common Shift	-.0567 (.0121)	0.0085 (.0152)
	Maximum Size Shift	-.0028 (.0129)	0.1054 (.0164)
Maximum Size - Summed Coefficients	Patented Brand	0.0495 (.0144)	0.0420 (.0170)
	Competitive Brand	0.0320 (.0175)	0.0789 (.0219)
	Off-Patent Brand	0.0417 (.0083)	0.0378 (.0094)
	ANDA holder	-0.0133 (.0238)	-.1033 (.0302)
	Generic Distributor	-.0567 (small)	0.1139 (.0204)

Table X: Quantity Shifts from December to March 1991
Full Dataset
Weighted by Revenue and Inverse of Number of Observations per Drug

		Drugstore	Hospital
Patented Brand	Common Shift	0.0333 (.0427)	-.0247 (.0372)
	Maximum Size Shift	0.0189 (.0384)	-.1770 (.0328)
Competitive Brand	Common Shift	-.2162 (.0377)	-.0333 (.0248)
	Maximum Size Shift	-.2374 (.0238)	-.0504 (.0226)
Off-Patent Brand	Common Shift	0.1824 (.0266)	-.1444 (.0250)
	Maximum Size Shift	0.5053 (.0231)	0.1322 (.0210)
ANDA	Common Shift	-.1589 (.0494)	0.0096 (.0608)
	Maximum Size Shift	0.2974 (.0393)	0.0752 (.0396)
Generic	Common Shift	-.1309 (.0947)	0.0044 (.0688)
	Maximum Size Shift	0.8886 (.0462)	0.2542 (.0494)
Maximum Size - Summed Coefficients	Patented Brand	0.0522 (.0330)	-.2017 (.0256)
	Competitive Brand	-.4536 (.0354)	-.0837 (.0234)
	Off-Patent Brand	0.6877 (.0166)	-.0122 (.0187)
	ANDA holder	0.1385 (.0540)	0.0848 (.0635)
	Generic Distributor	0.2996 (.0811)	0.2586 (.0745)

Table XI: Quantity Shifts from December 1990 to March 1991
Balanced Dataset
Weighted by Revenue and Inverse of Number of Observations per Drug

		Drugstore	Hospital
Patented Brand	Common Shift	0.0364 (.0566)	-0.0174 (056)
	Maximum Size Shift	0.0375 (.052)	0.0688 (.051)
Competitive Brand	Common Shift	-0.0660 (.0442)	-0.1257 (.044)
	Maximum Size Shift	-0.1726 (.028)	0.1406 (.027)
Off-Patent Brand	Common Shift	-0.1284 (.028)	-0.1496 (.028)
	Maximum Size Shift	0.2432 (.030)	0.1551 (.029)
ANDA	Common Shift	-0.1091 (.055)	-0.1342 (.054)
	Maximum Size Shift	0.2902 (.034)	0.4001 (.033)
Generic	Common Shift	-.0613 (.067)	-0.4066 (.067)
	Maximum Size Shift	0.0713 (.039)	0.5092 (.038)
Maximum Size - Summed Coefficients	Patented Brand	0.0739 (.043)	0.0515 (.043)
	Competitive Brand	-.2386 (.044)	0.0149 (.043)
	Off-Patent Brand	0.1148 (.020)	0.0054 (.020)
	ANDA holder	0.1811 (.058)	0.2659 (.057)
	Generic Distributor	0.0101 (.071)	0.1026 (.070)

Table XII: Quantity Shifts from December 1990 to March 1991
Big Firms Only
Balanced Dataset
Weighted by Revenue and Inverse of Number of Observations per Drug

		Drugstore	Hospital
Patented Brand	Common Shift	0.0169 (.097)	-0.1520 (.174)
	Maximum Size Shift	-0.0323 (.108)	0.5803 (.107)
Competitive Brand	Common Shift	-0.0253 (.082)	-0.0437 (.168)
	Maximum Size Shift	-0.2483 (.065)	-.0034 (.086)
Off-Patent Brand	Common Shift	-0.1346 (.059)	-0.2920 (.104)
	Maximum Size Shift	0.3056 (.065)	0.2936 (.109)
ANDA	Common Shift	-0.0756 (.062)	-0.1224 (.115)
	Maximum Size Shift	0.3467 (.115)	0.0603 (.319)
Generic	Common Shift	-0.5172 (.058)	-0.1437 (.602)
	Maximum Size Shift	0.8590 (.084)	0.1139 (.233)
Maximum Size - Summed Coefficients	Patented Brand	-.0553 (.0859)	0.4283 (.196)
	Competitive Brand	-.2177 (.1105)	-0.0471 (.188)
	Off-Patent Brand	0.1347 (.0478)	0.0016 (.108)
	ANDA holder	0.4591 (.1529)	-0.0621 (.334)
	Generic Distributor	0.5860 (.1029)	-0.0298 (.640)

Table XIII: ANDA Holder and Generic Distributor Price and Quantity Shifts around MFN Implementation Balanced Dataset Only Generic Firm Observations Included[82]

January to March 1991 Shift		Log Price			
		Drugstore		Hospital	
		ANDA	Generic	ANDA	Generic
Number of ANDAs < 4	Shift	0.012	-0.039	-0.018	0.029
		(0.024)	(0.022)	(0.021)	(0.019)
N = 2540	Maximum Shift	-0.009	-0.344	-0.036	0.104
		(0.030)	(0.028)	(0.026)	(0.024)
Number of ANDAs > 7	Shift	-0.441	-0.004	0.031	0.125
		(0.028)	(0.037)	(0.036)	(0.045)
N = 9705	Maximum Shift	-0.081	0.006	-0.105	-0.087
		(0.033)	(0.051)	(0.042)	(0.065)

February to March 1991 Shift		Log Quantity			
		Drugstore		Hospital	
		ANDA	Generic	ANDA	Generic
Number of ANDAs < 4	Shift	0.422	0.501	0.347	0.779
		(.085)	(0.085)	(.072)	(0.067)
N = 2540	Maximum Shift	-0.116	0.276	-0.095	-0.483
		(.105)	(0.093)	(.089)	(0.078)
Number of ANDAs > 7	Shift	0.259	-0.179	-0.448	-0.089
		(.047)	(0.053)	(.043)	(0.056)
N = 9705	Maximum Shift	-0.383	0.179	0.541	-0.034
		(.071)	(0.100)	(.074)	(0.105)

[82] Presentation intercepts are included in each regression, as are chemical-class time trends. Standard errors are in parentheses.

Table XIV: ANDA Holder and Generic Distributor Price and Quantity Shifts around MFN Implementation
Full Dataset
Only Generic Firm Observations Included[83]

January to March 1991 Shift		Log Price			
		Drugstore		Hospital	
		ANDA	Generic	ANDA	Generic
Number of ANDAs < 4	Shift	0.044	0.013	0.053	0.056
		(.014)	(0.017)	(.018)	(.016)
N = 7068	Maximum Shift	-0.038	-0.059	-0.047	0.088
		(.014)	(.011)	(.020)	(.017)
Number of ANDAs > 7	Shift	-0.365	0.058	0.097	0.102
		(.020)	(.022)	(.028)	(0.025)
N = 20833	Maximum Shift	0.472	0.093	-0.094	-0.001
		(.011)	(.022)	(.018)	(0.034)

February to March 1991 Shift		Log Quantity			
		Drugstore		Hospital	
		ANDA	Generic	ANDA	Generic
Number of ANDAs < 4	Shift	0.098	-1.40	0.124	-0.420
		(.085)	(.083)	(.039)	(.048)
N = 2734	Maximum Shift	0.124	2.246	-1.21	0.349
		(.095)	(.073)	(.095)	(0.068)
Number of ANDAs > 7	Shift	0.120	0.630	-0.238	-0.014
		(.078)	(.061)	(.048)	(.044)
N = 11828	Maximum Shift	0.049	-0.379	0.125	-0.092
		(.089)	(.076)	(.045)	(.084)

[83] Presentation intercepts are included in each regression, as are chemical-class time trends. Standard errors are in parentheses. All adjusted R-squared values are greater than 0.90.