

Strategic Entry Deterrence and the Behavior of Pharmaceutical Incumbents Prior to Patent Expiration¹

Glenn Ellison
MIT and NBER

and

Sara Fisher Ellison
MIT

April 2007

¹email: gellison@mit.edu, sellison@mit.edu. This work was supported by the National Science Foundation (grants SES-9818534, SES-0219205, and SES-0550897), the Sloan Foundation, and the Hoover Institute. We thank Emek Basker, Tom Chang, David Hwang, and Alan Sorensen for excellent research assistance, Gillian Currie for help in constructing the data set, Katie MacFarlane for guidance on institutional details, and Ditas Riad and Rhea Mihalison for their assistance on data issues. We also thank Fiona Scott Morton for her part in our joint data collection efforts and for providing other data to us. Finally, we thank Marcus Asplund, Ernie Berndt, Steve Berry, Richard Blundell, Judy Chevalier, Paul Joskow, Nancy Rose, Otto Toivanen, and seminar participants at various schools for helpful comments.

Abstract

This paper develops a new approach to testing for strategic entry deterrence and applies it to the behavior of pharmaceutical incumbents just before they lose patent protection. The approach involves looking at a cross-section of markets and examining whether behavior is nonmonotonic in the size of the market. Under certain conditions, investment levels will be monotone in market size if firms are not influenced by a desire to deter entry. Strategic investments, however, may be nonmonotone because entry deterrence is unnecessary in very small markets and impossible in very large ones, resulting in overall nonmonotonic investment. The pharmaceutical data contain advertising, product proliferation, and pricing information for a sample of drugs which lost patent protection between 1986 and 1992. Among the findings consistent with an entry deterrence motivation are that incumbents in markets of intermediate size have lower levels of advertising and are more likely to reduce advertising immediately prior to patent expiration.

Glenn Ellison
Department of Economics
Massachusetts Institute of Technology
50 Memorial Drive
Cambridge, MA 02142-1347

Sara Fisher Ellison
Department of Economics
Massachusetts Institute of Technology
50 Memorial Drive
Cambridge, MA 02142-1347

1 Introduction

The insight that firms may make “strategic investments” to alter future competitive conditions is one of the most fundamental ideas in industrial organization. The chapter on “Entry, Accomodation, and Exit” is easily the longest in Tirole’s (1988) text. In it, Tirole reviews arguments about how excess capacity, capital structure, advertising, contractual practices, learning-by-doing, and long-run decisions can be used to deter entry.¹ Strategic investment models are difficult to test directly, however, and the vast majority of this literature is theoretical. In this paper, we propose a new empirical approach for examining strategic entry deterrence.

Our applied focus is on the pharmaceutical industry. Using a panel of drugs that lost their U.S. patent protection between 1986 and 1992, we explore how pharmaceutical incumbents have dealt with the threat of generic entry. We examine incumbents’ advertising, product proliferation, and pricing decisions as patent expiration approached, and ask whether the behaviors appear to be influenced by an entry-deterrence motive.

We begin in Section 2 with a discussion of strategic entry deterrence and some motivation for our approach. We modify the textbook model to bring it closer to empirical applications: we assume that entry costs are random and unknown to the incumbent so that it is impossible to perfectly forecast whether entry will occur. We review what is meant by “strategic entry deterrence” in this setting. We note that the incentive to deter entry will be stronger in intermediate-sized markets than in very small or very large markets. In the former, no investments are needed to deter entry. In the latter, deterring entry is often impossible. A simple numerical example illustrates how the nonmonotonicity of the entry-deterrence incentive can lead to a nonmonotonic relationship between equilibrium investment levels and market size.

Our approach to testing whether firms are actively trying to deter entry is a classic reduced-form approach: we identify a prediction of the strategic investment model that differs depending on whether firms take entry-deterrence benefits into account when choosing

¹Some of the classic papers in this literature are Bulow, Geanakoplos, and Klemperer (1985), Fudenberg and Tirole (1983a, 1983b, 1984), Dixit (1980), Schmalensee (1978, 1981), Gelman and Salop (1983), Judd (1985), Aghion and Bolton (1987), and Cooper (1986).

their actions; and then test this prediction. The formal results underlying the approach are presented in Section 3. Our main IO theory result is a demonstration that, under specified conditions, actions will be monotonically related to a market size parameter *if firms are not influenced by an entry deterrence motive*. Several examples are used to provide intuition for the required conditions. The important implication of the theoretical result is that one can reject the null hypothesis that firms are not actively trying to deter entry by testing and rejecting the hypothesis that there is a monotonic relationship between firms’ actions and market size in a cross-section of markets.

Several recent papers have discussed ways of performing statistical hypothesis tests that a relationship is monotonically increasing.² Section 4 contains a brief discussion of this literature, a description of the tests we will use, and some additional monotonicity theorems relating to models with measurement error and endogenous right-hand side variables. Robustness to such factors is a potential advantage of an approach focusing on monotonicity.

In Section 5 we turn our attention to the pharmaceutical industry. The pharmaceutical industry is an important industry that has attracted a great deal of attention in policy circles. It also has several features that make it a nice environment in which to study strategic entry deterrence. One of these is that one can obtain a sizable sample of similarly situated incumbents facing a threat of entry by looking at manufacturers of branded drugs whose patent protection is about to expire. Another is that there are several potential tools that incumbents might use to deter entry, and we were able to obtain data on several of the most important: advertising, product proliferation, and pricing. A third is that there is a change in entry conditions within each market—entry is prohibited until a known date. Our approach to testing for strategic entry deterrence only requires a single cross-section, but having data both on actions immediately prior to patent expiration and actions in earlier years when patent expiration was less salient allows us to also implement a difference-in-differences version of our test.

The first thing we do in our analysis is to identify a proxy for “market size” and note that our dataset contains sufficient heterogeneity in market size to make it plausible that

²These include Hall and Heckman (2000), Ghosal, Sen and van der Vaart (2000), and Gijbels, Hall, Jones and Koch (2000).

we could find nonmonotonicities. Specifically, revenue received in the U.S. in the years immediately prior to patent expiration can serve as a proxy for market size because it is a strong predictor of whether generic entry will occur.³ The lowest-revenue drugs in our dataset are unlikely to ever face generic competition. For high-revenue drugs generic entry is a near certainty.

We then examine four incumbent behaviors that might plausibly be involved in an entry deterrence strategy. “Detail advertising” is the practice of sending representatives to doctors’ offices to promote a drug via one-on-one conversations. “Journal advertising” is the placement of advertisements for a drug in medical journals. We use the term “presentation proliferation” to refer to firms’ decisions about whether to sell a drug in small or large number of presentations, *e.g.* should it be offered just in a 100mg tablets or should the firm also produce a 50mg tablet, a 200mg tablet, a gelcap, an oral liquid, etc. Finally, we examine pricing. The literature on “limit pricing” is one of earliest and best known parts of the entry deterrence literature.⁴

We look for evidence that each of these behaviors is influenced by the entry-deterrence motive in two ways. First, we take a pure cross-sectional approach: we look at the relationship between each behavior and pre-expiration revenues in the cross-section of drugs and test whether the relationship is nonmonotonic. We find some evidence of nonmonotonicity in the journal advertising data. The form of the nonmonotonicity is that journal advertising is unusually low for drugs in intermediate-sized markets. This is what one would expect under a strategic-entry deterrence theory: firms in intermediate-sized markets have an incentive to let their market languish to make it less attractive to generic entrants.

Second, we examine how incumbents change their behavior as patent expiration nears. We find some evidence of a nonmonotonic relationship between behavior changes and market size in detail advertising and (less strongly) in product proliferation. The changes in detail advertising have a similar pattern to that noted above: it is most likely to be reduced in the intermediate-sized markets.

We conclude that there appears to be some evidence of strategic entry-deterrence by

³Grabowski and Vernon (1992), Bae (1997), and Scott Morton (2000) previously reported similar results.

⁴See, for example, Gaskins (1971), Milgrom and Roberts (1982), Fudenberg and Tirole (1986), and Klemperer (1987).

pharmaceutical incumbents. More broadly, we hope that our results also suggest that monotonicity tests may be a useful way to provide evidence on “strategic investment” theories in industrial organization and other fields.

Our paper can be seen as related to two empirical literatures in industrial organization. First, a number of papers have previously explored pricing, advertising, and entry in the pharmaceutical industry.⁵ Most closely related to our work is Caves, Whinston and Hurwitz (1991), a descriptive study based on thirty drugs with patents expiring between 1976 and 1987. They look mostly at the average behavior of incumbents before and after expiration and also separate drugs into low and high revenue categories and see if incumbent advertising behavior differs. Scott Morton (2000) focuses on the determinants of generic entry in a data set that overlaps substantially with ours. In addition to looking at exogenous market characteristics, she also looks for effects of incumbents’ advertising expenditures on the probability of generic entry. Grabowski and Vernon (1992) also study a panel of drugs with expiring patents and focus on post-entry behavior of both incumbents and generic entrants. Ellison and Wolfram (2006) examine pricing as a potentially strategic investment to forestall future regulation. They find that price increases by pharmaceutical firms during the Clinton health care reform debate were related to measures of firms’ potential losses from drug price regulation.

A second literature to which we contribute is the empirical literature on strategic entry deterrence (and entry accommodation). Developing structural tests of whether particular investments are strategic has been seen as difficult. The one paper we are aware of that has attempted this approach is Kadiyali’s (1996) study of the market for film. Kadiyali estimates price and advertising elasticities and argues that observed levels of price and advertising by Eastman Kodak were inconsistent with static monopoly profit maximization but consistent with entry deterring behavior. It has been more common to provide indirect evidence that investments are chosen strategically by showing that investments do affect

⁵Some notable papers are, for instance, Masson and Steiner (1985), Comanor (1986), Grabowski and Vernon (1990), Caves, Whinston and Hurwitz (1991), Frank and Salkever (1992), Scherer (1993), Berndt, Griliches and Rosett (1993), Griliches and Cockburn (1994), Berndt, Cockburn and Griliches (1996), Frank and Salkever (1997), Ellison, Cockburn, Griliches and Hausman (1997), Lu and Comanor (1998), Ellison (1998), Scott Morton (1999), and Ellison and Wolfram (2006). There are a number of books of interest about the economics of pharmaceuticals including Walker (1971), Schwartzman (1976), Temin (1980), and Schweitzer (1997).

future competition (which will lead us to conclude that investments must be strategic if we believe that firms are rational and aware of the effect on competition). One can think of Chevalier's (1995a, 1995b) studies of the effect of capital restructuring on entry and exit and supermarket pricing, Lieberman's (1987) discussion of the responses by incumbents in chemical industries to rivals' additions of capacity, and Scott Morton's (2000) discussion of the effects of advertising on entry as providing evidence of this sort. Lieberman also looks for evidence of entry deterring behavior in cross-sectional patterns in examining whether there is more excess capacity in markets which are more concentrated. A third approach taken by Goolsbee and Syverson (2004) in the airline industry is to examine how incumbent behavior changes in response to exogenous changes in potential entry that otherwise have no effect on current competitive conditions.⁶

2 Strategic entry deterrence

In this section we develop a simple model to review the idea of strategic entry deterrence and bring out its implications in a framework suited to empirical applications. We use a numerical example to illustrate how nonmonotonic patterns can arise in cross-section data.

2.1 A model

The prototypical model of strategic entry deterrence is a three-stage game like the first one in Figure 1. In the first stage, the incumbent firm 1 chooses an investment level A at a cost of $c(A)$. Assume that $c'(A) > 0$ and $c''(A) \geq 0$. Before the second stage, the potential entrant (firm 2) observes the incumbent's choice of A . Firm 2 then chooses whether to enter the market, which requires paying a sunk cost of E . In the third stage, either the incumbent is a monopolist or the incumbent and entrant compete as duopolists. If the incumbent is a monopolist, assume that it chooses some action $x_1^m(A)$ in the third period and as a result earns profits, $\pi_1^{m*}(A) \equiv \pi_1(x_1^m(A), A)$. If entry occurs, assume that the unique Nash equilibrium of the third stage game involves the firms choosing actions $x_1^*(A)$ and $x_2^*(A)$ and receiving profits $\pi_i^{d*}(A) \equiv \pi_i^d(x_1^*(A), x_2^*(A), A)$. Assume that $\pi_1^{m*}(A)$ and

⁶Other approaches have also been taken in a few papers. Smiley (1988) reports evidence from surveys of firms about what strategies they use to deter entry. Cooper, Garvin and Kagel (1997) examines a limit-pricing model experimentally. Dafny (2005) applies our approach in studying hospital markets.

$\pi_i^{d*}(A)$ are concave, and that the firms' best responses are always interior and given by the unique solution to the first-order conditions.

Tirole (1988) describes how a large number of classic papers in industrial organization (and corporate finance and international trade) can be fit within this framework. The key insight is that the “investment” A can be any action that is not easily undone. The could be a standard investment like building a new plant that will have a lower marginal production cost. It could also, however, be something like a lobbying expenditure, choosing a product design that makes the product less (or more) similar to other products, building up goodwill through an advertising campaign, or taking on debt. Entry deterring investments can be welfare-reducing, but this need not be the case.

The one departure we have made from the way strategic investment models are presented in Tirole (1988) (and in most papers) is that we assume that the entry cost E is stochastic with CDF $F(E)$ and that Firm 2 learns the entry cost before making its entry decision. The primary consequence is that Firm 1 will not know for sure whether entry will occur when making its investment decision. We think that this is a more realistic depiction of the situation firms face in the real-world and as well as a necessary modification for empirical applications.

2.2 The strategic entry deterrence incentive

In this model, the incumbent firm 1 is said to practice “strategic entry deterrence” in that it “distorts” A away from the level that maximizes profits (holding firm 2's entry probability fixed) in hopes of deterring entry by firm 2. More precisely, let A_{ED}^* be the sequential equilibrium choice of A in this model. What IO economists mean when they say that investment is “distorted” is that A_{ED}^* differs from the investment level, A_{ND}^* , that would be chosen in the second game pictured in Figure 1.⁷ In the second game, firm 2 does not observe firm 1's investment level until after the entry decision has been made. Hence, investment can not have a causal effect on the entry decision. The nonstrategic investment

⁷To avoid confusing people who know the literature we should note that we have simplified the standard presentation to omit any mention of strategic entry accomodation. Our assumption that A is observed at $t = 2\frac{1}{2}$ in the model “with no strategic entry deterrence motive” implies that both A_{ND}^* and A_{ED}^* reflect strategic entry accomodation effects. Hence, any differences between A_{ED}^* and A_{ND}^* are entirely due to entry deterrence motives.

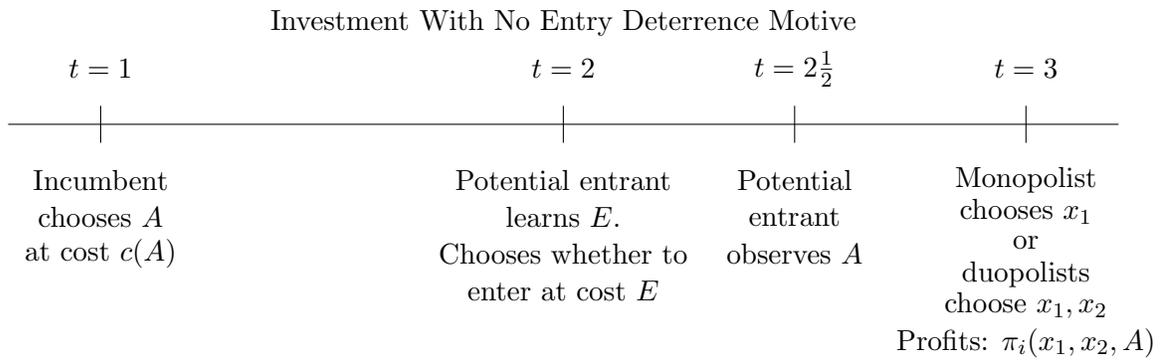
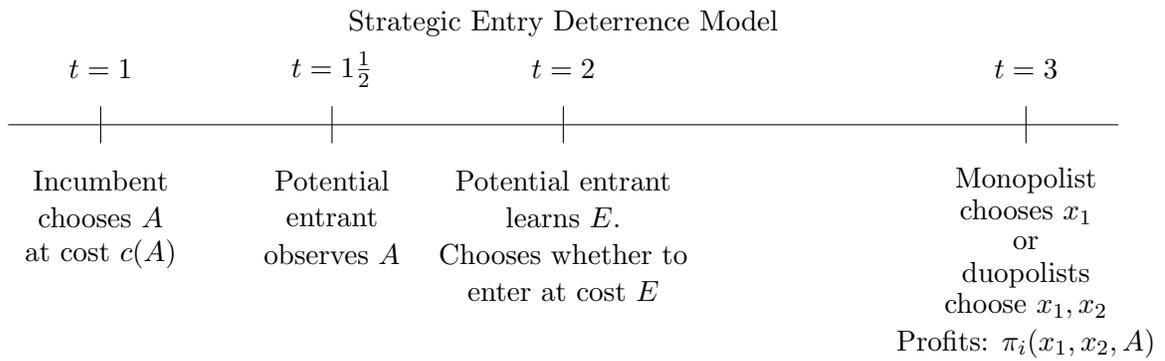


Figure 1: The model

level A_{ND}^* can be thought of either as reflecting what would happen if there was no entry deterrence motive or as reflecting what would happen if there was an entry deterrence motive but the incumbent ignored it. Whether incumbents behave in this manner can be of interest for several reasons. For example, antitrust authorities may insist that firms do not take actions that serve only to eliminate future competition, and economists may want to know whether firms are sufficiently rational and forward looking to invest strategically (and whether the long-run consequences of an investment are what we think they are).

Aspects of the strategic entry deterrence motive are most apparent in the first-order conditions that describe A_{ED}^* and A_{ND}^* . In the strategic entry deterrence model, firm 1's expected profit is a function of its first period investment:

$$E(\pi_1(A)) = F(\pi_2^{d^*}(A))\pi_1^{d^*}(A) + (1 - F(\pi_2^{d^*}(A)))\pi_1^{m^*}(A) - c(A).$$

In the model with no strategic entry deterrence motive, firm 1's expected profit depends both on the actual value of A and on firm 2's belief about the value of A that was chosen in the first period. In equilibrium, firm 2 will assign probability one to firm 1 having chosen A_{ND}^* , so firm 1's profit function is

$$E(\pi_1(A, A_{ND}^*)) = F(\pi_2^{d^*}(A_{ND}^*))\pi_1^{d^*}(A) + (1 - F(\pi_2^{d^*}(A_{ND}^*)))\pi_1^{m^*}(A) - c(A).$$

The first-order conditions for the equilibrium investment levels in the two models are thus

$$\begin{aligned} c'(A_{ND}^*) &= F(\pi_2^{d^*}(A_{ND}^*))\frac{\partial \pi_1^{d^*}}{\partial A}(A_{ND}^*) + (1 - F(\pi_2^{d^*}(A_{ND}^*)))\frac{\partial \pi_1^{m^*}}{\partial A}(A_{ND}^*) \\ c'(A_{ED}^*) &= F(\pi_2^{d^*}(A_{ED}^*))\frac{\partial \pi_1^{d^*}}{\partial A}(A_{ED}^*) + (1 - F(\pi_2^{d^*}(A_{ED}^*)))\frac{\partial \pi_1^{m^*}}{\partial A}(A_{ED}^*) \\ &\quad + (\pi_1^{d^*}(A_{ED}^*) - \pi_1^{m^*}(A_{ED}^*))\frac{d\pi_2^{d^*}}{dA}(A_{ED}^*)f(\pi_2^{d^*}(A_{ED}^*)). \end{aligned}$$

The difference between the two first order conditions is the presence of the final term in the first-order condition for A_{ED}^* . This term is the “strategic entry deterrence” incentive. Because firm 1's profit is higher when it is a monopolist, it has an incentive to distort its investment to reduce firm 2's profit (which reduces the likelihood of firm 2 entering).

The main observation about the strategic entry deterrence incentive we would like to highlight is that it may be larger in intermediate-sized markets than in very small or very

large markets. The incentive is a product of three terms. The third of these, $f(\pi_2^{d*}(A_{ED}^*))$, is the likelihood that firm 2's fixed entry costs are exactly equal to the equilibrium profits firm 2 would earn at the post-entry stage, making firm 2 indifferent between entering and not entering. In very small markets this likelihood will be small because the fixed entry costs will almost surely be much larger than the duopoly profits. In very large markets it will be small because the fixed entry costs will almost surely be much smaller than the duopoly profits. In intermediate-sized markets there is a greater chance that the investment will have a pivotal effect on entry.

2.3 An example of entry-deterrence in a cross-section of markets

In this section, we present a concrete example of a strategic investment model and discuss cross-sectional implications.

Example 1 *Consider a cross-section of markets. Suppose that the i^{th} market has a mass z_i of potential consumers, but that the markets are otherwise identical. Let A reflect expenditures on a form of advertising that raises potential consumers' valuations for all products in the product class. More specifically, assume that each market contains consumers with heterogeneous types, θ , distributed uniformly on $[0, 1]$, and that if the monopolist spends $z_i A^2/2$ on advertising in market i , a consumer of type θ receives utility $\theta A - p_1$ if he buys the (branded) good from firm 1 at price p_1 , $\frac{1}{2}\theta A - p_2$ if he buys the (generic) good from firm 2 at price p_2 , and zero if he buys neither good.*

In the final period of this model it is easy to check that a monopolist sets $p_1 = \frac{A}{2}$ and receives profit $\frac{zA}{4}$. The duopoly equilibrium prices are $p_1^* = \frac{2}{7}A$ and $p_2^* = \frac{1}{14}A$. Duopoly profits are $\frac{8}{49}zA$ and $\frac{1}{49}zA$.

Figure 2 contains a graph of the equilibrium advertising levels in this model when the distribution F of entry costs is log normal with mean 0.0025 and variance 0.0015. In the model without entry deterrence motives, A declines smoothly from $\frac{1}{4}$ at $z = 0$ to $\frac{8}{49}$ in the limit as $z \rightarrow \infty$.⁸ When there is also an entry deterrence motive, advertising levels are similar when z is small, but substantially lower in markets of small to intermediate size

⁸Note that in order to show what happens as z goes from zero to infinity we have rescaled the x-axis on the graph using $x = z/(z + 1)$.

as firm 1 distorts its advertising downward to deter entry. In larger markets firm 1 begins to give up on entry deterrence, and the advertising levels in the strategic entry deterrence model approach the equilibrium values of the model without entry deterrence.

A notable feature of this example is that the relationship between advertising and market size is monotonic in the model without the entry deterrence incentive and nonmonotonic in the model with the entry deterrence incentive. In the section that follows we discuss the generality of this observation and the possibility of basing tests of strategic intent on it.

2.4 An aside on structural estimation

Few empirical papers have attempted to provide evidence on strategic entry deterrence via structural methods. At first, one might think that this is surprising: in theory a structural test is as simple as testing which of the competing first-order conditions for A best fits the data. In practice, however, such an approach can be very difficult. By the very nature of these models, entry deterrence can only occur when investments have long term consequences, which is a case when econometric estimates are more difficult. Also, to compute the terms in the first order conditions, one needs estimates of the incumbent's prior on the likelihood of entry, and what the long run benefit of the investment would have been in the counterfactual state of the world in which entry did or did not occur. Each of these can also be a challenge to estimate.⁹

3 Some Results on Monotonicity and Entry Deterrence Motives

In the classic reduced-form approach to empirical industrial organization, one identifies where competing models make different predictions and then tests those differences. The null hypothesis for the test we have in mind is that investments are not influenced by the strategic entry deterrence motive. In this section we discuss conditions under which investments that are not influenced by the strategic entry deterrence motive will be monotone in the market size. Under those conditions, if the data are nonmonotone, one can conclude

⁹See Kadiyali (1996) for one structural estimation. Recent advances in the estimation of dynamic structural models, *e.g.* Aguirregabiria and Mira (2007), Bajari, Benkard, and Levin (2006), and Berry, Ostrovsky, and Pakes (2005), should provide additional opportunities for work along these lines.

that investments are influenced by the strategic entry deterrence motive (or that auxiliary assumptions of our propositions are violated).

3.1 A basic monotonicity result: the direct and competition effects

Consider the model of investment without an entry deterrence motivation. Suppose that the profit and cost functions also depend on a characteristic z of the market. Our leading example will be the number of potential consumers in the market. Assume that the variable z is ordered so that larger values of z correspond to markets that are more profitable for firm 2, *i.e.*, $\frac{\partial}{\partial z}\pi_2^{d*}(A, z) > 0$.

In the nonstrategic investment model, investments will covary with z for two reasons.

Definition 1 *The “direct effect” of z on A is $F(\pi_2^*)\frac{\partial^2\pi_1^{d*}}{\partial z\partial A} + (1 - F(\pi_2^*))\frac{\partial^2\pi_1^{m*}}{\partial z\partial A} - \frac{\partial^2c}{\partial z\partial A}$.*

The direct effect is positive if increasing z raises the marginal benefit from the investment more than it raises the marginal cost of the investment (holding entry probabilities fixed). When the direct effect is positive, it gives the incumbent an incentive to invest more when z is larger. A negative direct effect gives the opposite incentive.

Definition 2 *The “competition effect” of z on A is $\frac{\partial\pi_1^{d*}}{\partial A} - \frac{\partial\pi_1^{m*}}{\partial A}$.*

The competition effect is positive if the marginal benefit of the investment is larger when firm 1 is engaged in duopoly competition than it is when firm 1 is a monopolist. A larger value of z makes it more likely that firm 2 will enter. When the competition effect is positive, it provides an incentive for firm 1 to invest more when z is larger.

The following simple proposition identifies a set of circumstances in which investment levels will be monotone in z .

Proposition 1 *Let $A_{ND}^*(z)$ be the equilibrium investment level in the model of investment absent entry deterrence motivations described above. Suppose $\frac{d\pi_2^{d*}}{dz} > 0$.¹⁰ Then $A_{ND}^*(z)$*

¹⁰Note that this does involve an additional assumption. We had earlier assumed just that z was ordered so that $\frac{\partial\pi_2^{d*}}{\partial z} > 0$. Because $\frac{d\pi_2^{d*}}{dz} = \frac{\partial\pi_2^{d*}}{\partial z} + \frac{\partial\pi_2^{d*}}{\partial A}\frac{dA}{dz}$, the added assumption can be thought of as a requirement that the direct effect of z on firm 2’s profits is greater than the indirect effect that comes from firm 1 changing its investment level in response to changing market conditions. While this assumption is often satisfied, it is stronger than is necessary. By expanding $\frac{d\pi_2^{d*}}{dz}$ before solving for $\frac{dA}{dz}$ it is easy to see that it suffices to

is monotone increasing if the direct and competition effects are positive and $A_{ND}^*(z)$ is monotone decreasing if the direct and competition effects are negative.¹¹

The proof of this proposition is given in the Appendix.

REMARK 1: Proposition 1 is not a result that says that investment without an entry deterrence motive is monotone in z provided some minor technical conclusions hold. We get monotonicity *if* the direct and competition effects work in the same direction. In some applications, the two effects go in the same direction. In others they would not. One must think about whether an application is of the former or latter type before interpreting a violation of monotonicity as evidence that investments reflect strategic entry deterrence concerns.

REMARK 2: When examining the relationship between investment levels and market sizes, one has substantial latitude in defining the left- and right-hand-side variables. Monotone transformations of either variable (like taking the log of A) will not affect whether a relationship is monotonic, but other choices one makes in defining variables clearly can. For example, a graph of total advertising expenditure vs. population could look very different from a graph of per capita advertising expenditure vs. population. Appropriate choices for variable definitions will reflect two concerns: the variables should be chosen so that one would expect the direct and competition effects to be of the same sign; and one should define them so that the direct effects are not so strong so as to make it implausible that strategic entry deterrence motives will be strong enough to lead to nonmonotonicities. For example, in many applications total advertising expenditures will be approximately proportional

instead add the assumption that

$$\frac{\partial^2 c}{\partial A^2} - F(\pi_2^*) \frac{\partial^2 \pi_1^{d*}}{\partial A^2} - (1 - F(\pi_2^*)) \frac{\partial^2 \pi_1^{m*}}{\partial A^2} > f(\pi_2^*) \frac{\partial \pi_2^{d*}}{\partial A} \left(\frac{\partial \pi_1^{d*}}{\partial A} - \frac{\partial \pi_1^{m*}}{\partial A} \right).$$

This will always hold if the direction in which firm 1 changes A as competition becomes more likely reduces firm 2's profits (so that the right hand side is negative). For example, this would be the case for an investment in a form of nonrivalrous advertising which raised consumer awareness of or valuation for all products in a product class. Otherwise, it will be necessary that the term on the right hand side not be too large, which will hold, for example, if the distribution of entry costs is sufficiently diffuse so that the density term is sufficiently small.

¹¹To make the propositions easier to read, we have written them using words like increasing and positive rather than nondecreasing and nonnegative. The results extend in all of the obvious ways, *e.g.* investment is monotone nondecreasing if the direct and competition effects are both nonnegative, and investment is monotone increasing if the direct and competition effects are both nonnegative and one is strictly positive.

to the market size and strategic entry deterrence motives would need to be tremendously strong to make total advertising expenditures decreasing in market size over any range. It is more plausible that there might be nonmonotonicities in per capita expenditures.

REMARK 3: The proposition says nothing about what happens if the direct and competition effects go in opposite directions. Investment absent entry deterrence motives can be nonmonotone in z in such applications, but there is no strong reason to think this will usually be the case, *e.g.* it is not true that one effect tends to dominate at small z and the other at large z . It would not be unreasonable to look for evidence of strategic entry deterrence by looking for nonmonotonicities even in the absence of a convincing argument that the direct and competition effects went in the same direction. One would, however, need to recognize that one could fashion an alternate explanation for the nonmonotonicity by having different nonstrategic effects dominate in different regions, and one would want to consider whether such an explanation is plausible.

One noteworthy special case in which the monotonicity argument is particularly simple is when z is the number of potential consumers in the market and the profit and cost functions are directly proportional to z . In this case we have the following corollary.

Corollary 1 *In the model above, suppose $c(A, z) = zc(A, 1)$ and $\pi_i^{j*}(A, z) = z\pi_i^{j*}(A, 1)$ for $i = 1, 2$ and $j = d, m$. Then, the direct effect is zero. Hence $A_{ND}^*(z)$ will be monotone increasing if the competition effect is always positive and $A_{ND}^*(z)$ will be monotone decreasing if the competition effect is always negative.*

3.2 Examples of direct and competition effects

In this section we discuss the direct and competition effects in a few examples. The examples are related to our pharmaceutical application, and are also intended to help build intuition for the two effects.

We already saw one numerical example of a situation in which investment absent entry deterrence motives was monotone decreasing in the market size. Why the pattern of investment is as it is in this example is straightforward: The profit and cost functions in example 1 satisfy the hypotheses of Corollary 1. The competition effect is negative because the

benefit of the advertising is greater for a monopolist than for a duopolist. (The duopolist gains less both because advertising has a smaller impact on consumers' incremental preference for the incumbent's product relative to the entrant's and because the incumbent has a smaller market share in duopoly.)

Another useful example to think about is advertising that increases perceived differentiation between the incumbent's product and the entrant's. With proportional advertising costs we would again expect advertising levels in such models to be monotone. This time, however, we would expect advertising to be monotone increasing: differentiating advertising is more valuable to a duopolist (which relies on differentiation to maintain markups) than to a monopolist. Here is formal version:

Example 2 *Differentiating advertising.*

Consider a mass z of potential consumers with unit demands differentiated by a taste parameter θ which is uniformly distributed on $[-1, 1]$. Suppose that after firm 1 spends $zA^2/2$ on advertising, a consumer of type θ receives utility $1 - p_1$ if he buys the good from firm 1 at price p_1 , $1 + \theta t(1 + A) - p_2$ if he buys from firm 2 at p_2 , and zero if he makes no purchase.

With this specification duopoly profits are proportional to $1 + A$ and monopoly profits are independent of A . Hence, the competition effect is positive. Again, there is no direct effect. Hence, $A_{ND}^*(z)$ will be monotone increasing.

In this example, we would get the same result with another reasonable specification for advertising costs.

Example 3 *Differentiating advertising via broadcast media.*

Suppose that A represented total advertising expenditure in a model where the advertising technology (perhaps like direct-to-consumer television advertising) is such that all potential consumers see any ad, i.e., where $c(A, z) = c(A, 1)$ and $\pi_i^{j}(A, z) = z\pi_i^{j*}(A, 1)$.*

In this case the direct effect of z on A is positive so the direct and competition effects of differentiating advertising would work in the same direction.

Another type of investment we will think about in the pharmaceutical application is investments in product proliferation.

Example 4 *Product proliferation.*

Let z be the mass of potential consumers and let A be a fixed expenditure on developing new versions of the product to be located at various points of a horizontal taste space a la Hotelling. Suppose that costs of developing these new varieties are independent of size of the consumer population, $c(A, z) = c(A, 1)$. Suppose also that second stage monopoly profits are only slightly increasing in A (because idiosyncratic taste variation is small relative to the value of the good), whereas duopoly profits are more steeply increasing in A (because product proliferation leads firm 2 upon entry to choose to compete directly with only a subset of firm 1's products).

In this example, $A_{ND}^*(z)$ will again be monotone increasing. The direct effect is positive: the benefits of an investment A are increasing in z whereas the costs are independent of z . The competition effect is positive because the marginal value of the investment is greater for a duopolist.

4 Econometric issues

The theoretical results above illustrate a general idea: one could seek evidence that investment levels reflect a strategic entry-deterrence motive by first arguing that investment levels A should be monotone in a market size measure z absent entry-deterrence motives and then showing that they are actually nonmonotone. To develop this idea into one on which econometric tests can be based, the model obviously needs to be augmented to contain an error term that can account for measurement error in A , unobserved heterogeneity across markets, etc. In this section, we review the existing literature on monotonicity tests, describe the specific tests we will carry out on our data, and discuss additional econometric issues relevant when one moves from the theory to applications.

4.1 Standard tests of monotonicity

Several recent papers have proposed tests of the hypothesis that data $\{A_i, z_i\}$ are generated by a process

$$A_i = A^*(z_i) + \epsilon_i,$$

with $A^*(z)$ being monotone increasing in z and the ϵ_i being independent of z_i . These include Gijbels *et al.* (2000), Ghosal *et al.* (2000), and Hall and Heckman (2000). Hall and Heckman’s approach is simple and intuitive: if the true $A^*(z)$ is monotone increasing, then it is unlikely that there will be large ranges of z over which the relationship between A_i and z_i appears to be decreasing. This motivates forming a test statistic by looking at how strong of a downward relationship one can find by considering all ranges $R = [r_1, r_2]$ containing at least m datapoints. Specifically, they propose estimating a linear regression on the subset of the data with $z_i \in R$ for each such R and using the product of the regression coefficient $\hat{\beta}^R$ and the sample standard deviation of the z ’s in the range σ_z^R as a measure of the strength of any decreasing relationship, *i.e.* they set

$$T_{HH} = \max_{|R \cap \{z_1, \dots, z_n\}| \geq m} -\hat{\beta}^R \sigma_z^R.$$

They show that critical values can be obtained by a bootstrap with normal errors or by a nonparametric bootstrap provided that m increases sufficiently quickly in n .

4.2 Our implementation

In this paper we will test for monotonicity in two ways: one uses a slight modification of Hall and Heckman’s test statistic; the other uses a new statistic we propose.

Our modification of Hall and Heckman’s test statistic is necessitated by the fact that we want the null hypothesis to be that $A^*(z)$ is monotone rather than monotone increasing. To this end, we set

$$T_{HH} = \min\left\{ \max_{|R \cap \{z_1, \dots, z_n\}| \geq m} -\hat{\beta}^R \sigma_z^R, \max_{|R \cap \{z_1, \dots, z_n\}| \geq m} \hat{\beta}^R \sigma_z^R \right\}.$$

Intuitively, this will be large if there are both ranges over which the data are increasing and ranges over which the data are decreasing.

The second test statistic we try assesses how well the data can be fit by a monotone function. Specifically, we use isotone regression to determine the monotone function $\hat{f}(z)$ that best fits the data, form the residuals $\hat{\epsilon}_i \equiv A_i - \hat{f}(z_i)$, and use a test statistic like that in Ellison and Ellison (2000) to test whether the residuals appear to come from a misspecified model:

$$T_{EE} = \frac{\hat{\epsilon}' \bar{W} \hat{\epsilon}}{\sqrt{2} \hat{\sigma}^2 \sum_{ij} \bar{w}_{ij}^2},$$

where W is a kernel weight matrix reflecting differences in the z 's and $\bar{W} = (W + W')/2$. Intuitively, if the true $A^*(z)$ is nonmonotone, then there will be regions where $A^*(z) > \hat{f}(z)$ and other regions where $A^*(z) < \hat{f}(z)$. This test looks for such regions by looking at whether the residuals from nearby observations are positively correlated.

We obtained critical values for each of these tests via bootstrap methods. We have not tried to extend the existing results to obtain formal proofs that that this procedure is valid in our setting.¹² We, therefore, conduct simulations to help assess the validity of the procedure and the power of the tests. We discuss these briefly in the appendix.¹³

4.3 Measurement error

A second issue that will come up in many applications is that one may have only an imperfect proxy for “market size” z . For example, in a dataset examining a cross-section of cities or countries, one would typically use population as a proxy for market size, which would not allow for taste differences across markets.¹⁴ We note here that this is often not a problem for our approach.

Suppose z is unobserved, but the data contain a proxy r correlated with z . Given a dataset containing observations $\{A_i, r_i\}$ satisfying $A_i = A^*(z_i) + \epsilon_i$ and appropriate regularity conditions, one can estimate the function $A(r)$ defined by $A(r) \equiv E(A^*(z)|r)$. Whether looking for nonmonotonicity remains an appropriate way to look for evidence of strategic entry deterrence depends on whether $A(r)$ inherits the monotonicity of $A^*(z)$ under the null.

Obviously, we can only hope that $A(r)$ will be monotone in r if the joint distribution of r and z is such that higher values of r are associated with higher values of z . It is a standard result in incentive theory that the appropriate meaning of association for many such problems is having the monotone likelihood ratio property (MLRP). The conditional

¹²The theoretical results of Hall and Heckman (2000) cannot be applied directly for a couple reasons: we have modified the test statistic to make it two-sided, and in some of our applications the A variable is discrete, which does not fit with their assumption that the ϵ_i are i.i.d. The theoretical results in Ellison and Ellison (2000) are inapplicable because it is assumed there that the function $\hat{f}(z)$ is obtained via a parametric method, whereas the $\hat{f}(z)$ in this paper is a nonparametric estimate.

¹³In unreported simulations we also examined the power of the Ghosal *et al.* (2000) and Gijbels *et al.* (2000) tests. We chose to use the two statistics we use because they were quicker to compute and/or more powerful in these simulations.

¹⁴See Bronnenberg, Dhar, and Dube (2005) for an interesting study of cross-market taste differences.

density $f(x|\theta)$ of a random variable is said to have the MLRP in x if $f(x|\bar{\theta})/f(x|\underline{\theta})$ is monotone increasing in x whenever $\bar{\theta} > \underline{\theta}$. Under this assumption we have

Proposition 2 *Suppose $A^*(z)$ is monotone in z . Suppose distribution of r conditional on z has the MLRP in r . Then, $A(r)$ is monotone in r .*

Proof

MLRP implies that the distribution of z conditional on r is increasing in the first order stochastic dominance sense. (See Milgrom (1981).) This implies that the expectation conditional on r of any increasing function of z is increasing in r . $A(r)$ is the expectation of $A^*(z)$.

QED.

In the classic measurement error model, $r_i = z_i + \epsilon_i$, the MLRP holds provided that the density g of ϵ has $g(\epsilon - \delta)/g(\epsilon)$ increasing in ϵ for any $\delta > 0$. This property, called monotone likelihood ratio, holds for most standard distributions including the normal. We therefore think of Proposition 2 as indicating that measurement error is not a significant obstacle to our approach.¹⁵

4.4 Endogeneity

In some applications, one may also worry that the available proxies for market size are endogenous. For example, in our pharmaceutical application, our best proxy for the potential market for a drug is the total revenue that the incumbent monopolist has been receiving prior to patent expiration. This will be correlated with the number of potential consumers and their aggregate willingness to pay for the drug, but one would also expect that it would be influenced by the investments, *e.g.* revenue will be higher if the incumbent advertises more.

Endogeneity is a more serious concern than measurement error. However, because we are only interested in monotonicity (as opposed to obtaining consistent parameter estimates) endogeneity can be less of a problem for our approach than it would be for structural

¹⁵The textbook example of a situation in which the property fails is when the distribution of ϵ_i is extremely bimodal. For example, if z is standard normal and ϵ is always either -10 or 10, then observing $r = -9$ likely indicates that $z = 1$, whereas observing $r = 9$ likely indicates that $z = -1$.

estimation. One can easily envision two situations in which the use of an endogenous proxy without an instrument would not be a problem. First, if the endogeneity bias in estimating $\partial A/\partial z$ is of the same sign as $\partial A/\partial z$, then investments will still be estimated to be monotone in the endogenous variable under the null. The power of our test may be reduced, but a rejection of monotonicity would still be evidence of strategic entry deterrence motives. Second, if the endogeneity bias is of the opposite sign as $\partial A/\partial z$ but is sufficiently small, then the estimated investment function would also still be monotone under the null.

Formally, consider again a cross-section dataset containing investment levels A_i and a proxy r_i for the market size. Suppose the data generating process is

$$\begin{aligned} A_i &= A^*(z_i) + \epsilon_i \\ r_i &= r(z_i, A_i) + \eta_i, \end{aligned}$$

where z_i , ϵ_i , and η_i are unobserved independent random variables and $r(z, A)$ is a function that is monotone increasing in both arguments. Again, the function one can hope to estimate from the data is $A(r) \equiv E(A^*(z) + \epsilon|r)$.

Define $\tilde{r}(z, \epsilon) \equiv r(z, A^*(z) + \epsilon)$. One result showing that endogeneity need not be a problem is straightforward.

Proposition 3 *Suppose $A^*(z)$ is monotone increasing. Suppose the distribution of η has a monotone likelihood ratio, and that the distributions of $\tilde{r}(z, \epsilon)$ conditional on z and ϵ both have the MLRP in \tilde{r} . Then, $A(r)$ is monotone increasing in r .*

Proof

When η has a monotone likelihood ratio, the distribution of $\tilde{r}(z, \epsilon)$ is increasing in r in the sense of first order stochastic dominance (FOSD). When the distributions of $\tilde{r}(z, \epsilon)$ conditioned on each argument both have the MLRP, this in turn implies that the distributions of z and ϵ are both increasing in r in the FOSD sense. When $A^*(z)$ is monotone increasing, this implies that the distribution of $A^*(z) + \epsilon$ is increasing in r in the FOSD sense, which implies that $A(r)$ is increasing.

QED.

Proposition 3 has two primary limitations. The most important limitation for applied work is that it only covers the case in which $A^*(z)$ is monotone increasing, not to the case when $A^*(z)$ is monotone decreasing.

A second limitation is that we have assumed that $\tilde{r}(z, \epsilon)$ has the MLRP in both arguments. If $r(z, A)$ is linear in its arguments, then one of these assumptions is simply that the distribution of ϵ has a monotone likelihood ratio. The other is that the distribution of $z + \beta A^*(z)$ has a monotone likelihood ratio. Even if z is normally distributed, one could find monotone functions $A^*(z)$ for which $z + \beta A^*(z)$ has a bimodal distribution (choose a function with two broad flat portions separated by a steeper portion). This seems less important as a practical concern.

If $A^*(z)$ is monotone decreasing the argument in the proof of Proposition 3 goes through until the very last step.¹⁶ It does not go all the way through because we end up with $E(A^*(z)|r)$ decreasing in r and $E(\epsilon|r)$ increasing in r , leaving the sum $E(A^*(z) + \epsilon|r)$ indeterminate. In some applications with $A^*(z)$ monotone decreasing it will still be appropriate to interpret a test of whether $A^*(r)$ is nonmonotonic as a test for strategic entry deterrence. This would be justified in applications for which the $A^*(z)$ decreasing effect should outweigh the $E(\epsilon|r)$ increasing effect. The former effect will be stronger than the latter when $r(z, A)$ is more dependent on z than A , when the variance of z is large relative to the variance in ϵ , and when $A^*(z)$ is more steeply downward sloping.

5 The Pharmaceutical Industry

In this section we provide some background on the U.S. pharmaceutical industry, discuss strategic instruments that firms might try to use to deter generic entry, describe our dataset, and note that the dataset has the type of heterogeneity in “market size” required for our approach.

¹⁶The assumption that $\tilde{r}(z, \epsilon)$ has the MLRP in both arguments may also be more difficult to justify. For larger values of \tilde{r} to be associated with larger values of z we will need $d\tilde{r}/dz$ to be positive, which requires that the direct effect dr/dz is larger than the indirect effect through investment levels $dr/dA dA^*/dz$.

5.1 Industry background

Prior to 1984 all but the most popular drugs tended to retain their monopoly position in the U.S. market long after their patent protection expired. FDA regulations required any firm wanting to produce a generic substitute to repeat the lengthy process of tests and clinical trials to which the incumbent had been subjected. Things changed dramatically in the mid-1980's: the Waxman-Hatch Act (1984) reduced regulatory barriers to generic entry; and state laws mandating/allowing generic substitution by pharmacists boosted the market share of generic drugs.¹⁷

When a blockbuster drug like Prozac loses patent protection, generic entry is swift and sure—within 18 months Prozac faced 21 generic competitors and had lost more than 80% of its market. Most drugs, however, are not blockbusters. Many FDA-approved drugs never achieve much commercial success. Others have been largely supplanted by the time they lose protection. In this paper, we study a set of 63 drugs that lost patent or FDA exclusivity protection between 1986 and 1992.¹⁸ Whereas Prozac had annual revenues of about \$3 billion prior to patent expiration, the median drug in our sample only had annual revenue of around \$20 million. For such a drug, generic entry is much less certain: only thirty seven of the sixty three drugs in our sample faced generic competitors within three years of the expiration of their patents.

There are a number of “investments” that one could imagine pharmaceutical incumbent distorting in order to deter entry. The most obvious is advertising, which plays an extremely important role in pharmaceutical markets—an oft-cited statistic by critics of the

¹⁷See Grabowski and Vernon (1996).

¹⁸These drugs are a subset of those used in Scott Morton (2000). The sample is intended to be an as-complete-as-possible list of the drugs that lost patent protection in this period, although we were conservative in constructing the sample and only included drugs when we were sufficiently confident about the identification of the relevant patent and exclusivity restrictions. This can be difficult even though drug manufacturers are required by the FDA to report all relevant patents with expiration dates and the FDA publishes this information in the *Approved Product List* (“The Orange Book”). The reason is that which patents are truly relevant is not something the FDA can sort out, and it is clearly in the interests of the manufacturers to list patents even if their relevance is questionable. For the high revenue drugs, potential entry dates are often listed in trade publications and are, therefore, fairly easy to track down, absent court battles over expiration. Information is more difficult to come by for the smaller revenue drugs because potential entry into those drugs is usually not an important event. For those we relied more on FDA publications. Additional sources we used were lists of patent expiration dates published by the Generic Pharmaceutical Industry Association and Arthur D. Little, Caves, Whinston, and Hurwitz (1991), lists of ANDAs, and information on generics being produced in various issues of *Drug Facts and Comparisons*.

pharmaceutical industry is that more money is spent by the industry on marketing than on research and development.¹⁹ In the period we study, there were two main advertising channels: “detail” and journal advertising.²⁰ Detail advertising is the practice of having sales representatives visit doctors’ offices to inform them about studies assessing a drug’s effectiveness and otherwise promote the product in one-on-one conversations. Journal advertising means placing advertisements in medical journals and other publications read by doctors. Expenditures on detail advertising are typically much larger than expenditures on journal advertising. An important aspect of both of these advertising technologies is that increases in demand they generate will in part be captured by generic competitors: doctors will know that generic versions of a drug can provide the same benefits; and pharmacists may dispense a generic even when the doctor has written a prescription for a brand name.

A second potential instrument for strategic entry deterrence that has received much less attention is presentation proliferation. Many prescription drugs are sold in a large number of “presentations”. The tranquilizer Haldol, for example, is sold in $\frac{1}{2}$, 1, 2, 5, 10 and 20 milligram tablets, as a concentrated liquid in bottles, and as a solution for intravenous use in vials, ampules and disposable syringes. Increasing the variety of presentations could be a strategy for deterring entry. It makes it more costly for an entrant to reproduce the incumbent’s complete product line, and reduces the revenues of an entrant that avoids the cost increase by only producing some presentations.

An additional instrument that firms might use to deter entry is pricing. The theoretical literature has identified a number of ways in which pricing decisions may affect subsequent entry: prices may signal something about the incumbent or the market to the entrant, they may be distorted for signal jamming reasons, or there may be some more direct link between periods due to switching costs, learning by doing, etc.²¹ In the pharmaceutical

¹⁹See U.S. Congress, Office of Technology Assessment (1993). Bill Clinton, in a 1993 speech, said that the pharmaceutical industry spends “\$1 billion more each year on advertising and lobbying than it does on developing new and better drugs We cannot have profits at the expense of our children.” (*The Wall Street Journal*, February 16, 1993) Schweitzer (1997) provides a thorough discussion of research and marketing expenditures of pharmaceutical firms. See Caves, Whinston, and Hurwitz (1991), Hurwitz and Caves (1988), Leffler (1981), Berndt, Bui, Reiley, and Urban (1995) and Rizzo (1999) for discussions of other aspects of pharmaceutical advertising.

²⁰Direct-to-consumer advertising via mass media did not begin in earnest until the mid 1990’s. See Rosenthal *et al.* (2002) for a description of the practice and some documentation of its prevalence.

²¹Among the early papers in this literature are Milgrom and Roberts (1982), Fudenberg and Tirole (1983b),

industry it is plausible that generic entrants have some uncertainty about the slope of the demand curve, and also plausible that prices may have some commitment value due to the scrutiny directed at price increases. Hence, several of the mechanisms by which prices could be a strategic investment seem potentially relevant.

5.2 Data

Our basic data set includes 63 distinct chemical compounds that faced potential generic entry as the result of a patent or FDA exclusivity expiration between 1986 and 1992.²²

We collected the data on revenues, prices, and advertising from historical IMS audits of the pharmaceutical industry. Like all IMS sales data, the prices and revenues are those paid by the retail or hospital sector, in other words, essentially at the wholesale level. Our revenue data contain annual presentation-level wholesale revenues for all presentations of each drug in both the hospital and drugstore submarkets for five years: three years prior to patent expiration, the year of patent expiration and the year following patent expiration. We construct two variables from this data which we use to help measure the attractiveness of the market to potential entrants: *Revenue3* is the average annual revenue (in thousands of dollars) from hospital and drugstore sales in the three calendar years before but not including the year of patent expiration; and *HospFrac* is the fraction of total revenues in the calendar year prior to patent expiration which were due to hospital sales. All prices and revenues are in constant 1982-1984 dollars.

Our advertising data on each drug consist of two variables, *Detail* and *Journal*. The former is the number of minutes that pharmaceutical company “detailers” spent promoting the drug in direct conversations with physicians. The latter is an estimate of dollars spent on journal advertisements promoting the drug based on audits of medical journals. The advertising data is at a monthly frequency and includes 48 observations per drug covering the thirty six months prior to patent expiration, the month of patent expiration and the eleven subsequent months.

1986), Harrington (1986), and Klemperer (1987).

²²Our data set contains 71 drugs, where a drug is defined as a brand-name product sold by the patent-holder or licensee prior to expiration. Seven of our chemical compounds were sold under multiple brand-names, accounting for the discrepancy.

Our primary measure of the degree to which an incumbent has engaged in presentation proliferation, *PresHerf*, is a Herfindahl-style measure that is also constructed from the presentation-level revenue data. Specifically, we define $PresHerf_{it} = w_i \sum_k z_{idkt}^2 + (1 - w_i) \sum_k z_{ihkt}^2$, where w_i is the fraction of the sales of drug i which are made through drugstores and z_{idkt} and z_{ihkt} are the fractions of drug i 's revenues in year t in the drugstore and hospital markets, respectively, which are accounted for by presentation k .²³ *PresHerf* will be large in markets where a small number of presentations account for most of the revenues and smaller in markets where sales are more evenly divided among a larger number of presentations.

Because of the different presentations, a drug's price is difficult to define. (Prices for different presentations are clearly not set to equalize the total cost of a duration of treatment or in proportion to the quantity of the active ingredient.) In our study of pricing patterns, we look at changes in the drugstore and hospital prices of each drug using variables, *HPrice* and *DPrice*, which give the price of one particular presentation of each drug in the five year window around the year of patent expiration.²⁴

We obtained information on drug characteristics and whether generic entry did in fact occur from several other sources.²⁵ The primary variable we will use to study entry, *Entry3Yr*, is a dummy variable equal to one if at least one firm had an Abbreviated New Drug Application (ANDA) approved (allowing it to produce a generic version of the drug) within three years of the date at which a patent expires.²⁶ *Chronic* is set to zero for drugs

²³Defining presentations by differences at the wholesale level will in some cases be a poor reflection of how proliferation affects the costs of entry. For example, 100mg tablets sold to pharmacies in a 100 tablet bottle will be treated as different from 100mg tablets sold in bubble packs and as different from 100mg tablets sold to pharmacies in a 500 tablet bottle. The descriptors in our data at times do not make it clear how similar/different wholesale presentations are, but it did not appear that problems like those described above are very important in the aggregate. We would also have preferred to sum the presentation-by-presentation revenues across hospitals and drugstores before computing the sum of squares, but given the form of our data, this would have entailed a laborious manual matching. Given that 70 percent of the drugs have at least 90 percent of their sales in one submarket or the other, we felt that just taking weighted averages was a reasonable compromise.

²⁴We usually chose the presentation that had the highest revenue in the first year of our data.

²⁵These include *Drug Facts and Comparisons*, *Physician's Desk Reference*, the FDA's *Approved Drug Products with Therapeutic Equivalence Evaluations*, and discussions with physicians.

²⁶Caves, Whinston and Hurwicz (1991) and Scott Morton (1999) note that entry in pharmaceutical markets often does not occur immediately upon patent expiration, and that only part of the delay in attributable to uncertainties in the length of time necessary for ANDA approval.

which treat an acute condition and to one for drugs which treat a chronic condition.²⁷ *Psych* is an indicator for whether the drug primarily treats a psychological condition. *Topical* is an indicator for whether the drug is usually applied topically. *TherSubs* is the number of other chemical compounds in a drug’s therapeutic class, where we used therapeutic categories defined by *Drug Facts and Comparisons*. *Specialist* is a proxy for the extent to which the drug tends to be prescribed by specialists. It is obtained by computing the GINI coefficient for each therapeutic class of drugs from a table of frequency of prescription by various specialties. Cardiovascular drugs, for instance, have a GINI coefficient of 0.18 whereas ophthalmic drugs have a value of 0.35, indicating that prescriptions for cardiovasculars are more spread out across specialties than are prescriptions for ophthalmics. (Those two categories represent the minimum and maximum values.) Each drug is categorized in therapeutic class and assigned the GINI coefficient for its therapeutic class as its value of *Specialist*.

Summary statistics are presented in Table 3. The first five variables in the table are defined at the drug level. The mean of the *Entry3Yr* variable reflects that 37 of the 63 drugs experienced generic entry in the three year window. The mean of the *Revenue3* variable indicates that the average drug had annual revenues of \$39.4 million.

Detail3 and *Journal3* are average annual values of the advertising variables over the same three-year pre-expiration period for which *Revenue3* was computed. The values for the mean advertising ratios in Table 3 indicate that 1.4% of sales were spent on journal advertising and approximately 5% on detail advertising.²⁸ *PresHerf3* is an average of the *PresHerf* variable over the same three year period.²⁹ Although the average number of presentations per drug is greater than six, the mean value of 0.54 indicates that one or two presentations usually account for a large portion of revenues.³⁰ The *Detail3*, *Journal3*, and *PresHerf3* variables have 69 or 70 observations rather than 63 because we have defined

²⁷The variable is set to one-half for a few drugs which were judged to be intermediate on this dimension.

²⁸Our detailing data are in minutes. This calculation assumes a cost of \$10 a minute.

²⁹For seven of the drugs we are missing data for one of the three years. In these cases, the average was taken over the two years for which data was available.

³⁰Recall that the scale of a Herfindahl index is such that the index would be equal to $1/n$ if a drug is sold in n presentations and each receives equal revenues. Ten of our drugs are sold in a single presentation and hence have *PresHerf3* equal to one. The mean of *PresHerf3* for the remaining drugs is still 0.46.

them at the level of the brand-name rather than at the level of the drug.³¹

The $DPrice_t$ and $HPrice_t$ variables are yearly observations of the price of one presentation of each drug deflated by the Consumer Price Index. The summary statistics indicate that the average price increases in the drugstore and hospital markets are 1% and 1.9% above the rate of inflation.

The data on the $HospFrac$ variable reflect that drugstore revenues are usually substantially larger than hospital revenues.

5.3 Revenue as a proxy for market size

Our approach to studying strategic investment requires that we have a proxy for “market size”. and that there be sufficient heterogeneity in this variable. In this section, we note that pre-expiration revenues should be a good market-size proxy.

Previous work by Grabowski and Vernon (1992), Bae (1987), and Scott Morton (2000) has established that pre-expiration revenues are a significant predictor of generic entry. This should not be surprising: there is tremendous variation in revenues across drugs and no obvious reason why the fixed costs of developing drugs should be comparably heterogeneous and correlated. The first column of Table 5 reports estimates from a probit regression of $Entry3Yr$ on $Revenue3$ to verify that such a relationship exists in our data as well. The second column adds several other covariates to the regression. None except $Revenue3$ have a statistically significant effect on the likelihood of generic entry.³²

One can only expect to be able to find a nonmonotonic investment pattern in the cross-section due to strategic investment if the heterogeneity in market sizes is sufficiently large so that a dataset contains markets where the likelihood of entry is small, intermediate, and

³¹Hence, the seven drugs that are sold under multiple brand names contribute multiple observations to these regressions. The small number of missing observations are due to cases where we judged the data to be unreliable. Drugs for which no detail or journal advertising was performed are included and coded as zeros.

³²The point estimates are that drugs treating chronic conditions and drugs sold mostly through hospitals were more likely to face generic entry, although neither estimate is significant even at the 10% level. We would also find such estimates a bit surprising as they do not conform with intuitive findings in the previous literature about where markups are greatest: Sorensen’s (2000) study of dispersion in retail drug prices in New York State indicates that drugs treating acute conditions have higher retail markups (and less dispersion), and Ellison and Snyder (2001) and others report that hospitals pay lower wholesale prices for antibiotics than do drugstores. Scott Morton (2000) does report that entry is significantly more likely for drugs treating chronic conditions and for drugs where the hospital share of sales is larger in her analysis of a larger dataset which overlaps substantially with ours.

large. To give some feel for the degree of heterogeneity in our data, we divide our sample into five revenue-based subsamples. Table 4 reports the range of revenues in each group and the fraction of drugs in each group that experienced entry within three years of patent expiration.³³ The main observation to be taken away from this table is that one can think of the lowest quintile, Q1, as containing drugs that face a low probability of generic entry, those in the next quintile, Q2, as having an intermediate probability of generic entry, and those in the top three quintiles, Q3, Q4 and Q5, as having a high probability of generic entry. The strategic entry-deterrence motive will vary continuously with revenues (and other unobservables), but one can roughly regard it as being most salient for drugs in the second-lowest revenue quintile.

6 Evidence on Strategic Investment in Pharmaceuticals

In this section we examine our data on detail advertising, journal advertising, presentation proliferation, and pricing for evidence that the incumbents' actions reflect a strategic entry deterrence motive. We will do this in two ways. First, in the most straightforward application of the framework developed in this paper we look for evidence of nonmonotonicity in the cross-section. Second, we conduct a similar analysis using behavior in the early part of our data (when entry deterrence was less salient) as an additional control variable.

6.1 Cross-sectional patterns in incumbent behavior

In our theory section we presented conditions under which investment would be monotone in market size if firms are not influenced by an entry-deterrence motive, and noted that a strong entry-deterrence motive might be expected to produce a nonmonotone pattern. We now discuss the potentially strategic actions one by one. In each case, we discuss what might be expected absent entry deterrence motives, how behavior would be distorted to deter entry, and what patterns are observed in our dataset. We find some evidence in favor of the entry-deterrence hypothesis in the journal advertising data.

³³Note that the five "quintiles" in this table contain uneven numbers of drugs. Recall that we are sometimes treating our sample as 63 drugs and sometimes as 69 brand names/drug combinations. We do the latter more frequently, and hence chose to define the quintiles to have the same number of brand names in each. Note that the quintiles are only being used to provide a feel for the data and our nonparametric monotonicity tests do not involve any arbitrary cutoffs.

6.1.1 Detail advertising

Detail advertising is extremely important in the development of the market for a prescription drug. Our cross-sectional analysis will examine advertising-to-sales ratios.³⁴ What patterns might be indicative of strategic entry deterrence? Absent entry-deterrence motives we would expect to see these ratios decline slightly with market size. The “competition effect” is that there are reduced long-run benefits to advertising high-revenue drugs approaching patent expiration (because a substantial share of future prescriptions will instead go to generics). There is little “direct effect” because the fact that detailing efforts are focused on doctors who write a substantial number of prescriptions for a condition means that they are not particularly less productive for drugs that treat less common conditions. In a strategic-entry-deterrence model the strategic effect is that an incumbent might reduce detail advertising levels in intermediate sized markets. This reduction in advertising would reduce the attractiveness of the market to potential entrants, and entry could thereby be deterred.

For a first look at how the detailing-to-sales ratio varies with revenues, we estimate the regression

$$\frac{Detail3_i}{Revenue3_i} = \beta_0 + \beta_1 \log(Revenue3_i) + \beta_2 (\log(Revenue3_i) - \bar{R})^2 + \beta_3 Specialist_i + \epsilon_i,$$

where \bar{R} is the mean of $\log(Revenue3)$. We include $Specialist_i$ because detailing is more cost-effective for drugs prescribed by specialists than for drugs prescribed by nonspecialists (each of whom will only have a small number of patients who could benefit from the drug).

Coefficient estimates are reported in the first column of Table 6. The regression does not provide any evidence for a nonmonotonic pattern. Indeed, it provides little evidence of any patterns at all. The R^2 of the regression is low. The coefficient estimates on the revenue variables and the coefficient estimate on the *Specialist* variable are all statistically insignificant.

Table 7 presents additional descriptive evidence and formal tests for nonmonotonicity. The first five columns of the first row of the table give the mean values of $Detail3/Revenue3$

³⁴For drugs sold under multiple brand names our advertising data is at the level of the brand rather than the drug. In these cases, we use brand-specific sales in the advertising to sales ratios.

within each of the revenue quintiles. The most noteworthy value is the low value for the second quintile, which, as noted earlier, can be thought of as drugs facing an intermediate probability of generic entry. To provide formal tests of nonmonotonicity we carried out the tests described in section 4.2. We used a parameter for the Hall-Heckman test corresponding to subregressions of fifteen observations. We set the smoothing parameter for the Ellison-Ellison style test to be equal to one-fifth of the support of the regressor. The sixth and seventh columns of the Table report estimated p-values for these two tests, neither of which rejects monotonicity at standard levels of significance. Apparently, the level of noise is sufficiently high and/or the pattern is sufficiently sensitive to the cutoffs so that the nonmonotonic pattern in the quintile means is not significant in our nonparametric tests.

6.1.2 Journal advertising

The second potential instrument for strategic entry deterrence available to us is journal advertising. What patterns might be expected here with and without strategic entry deterrence?

The direct effect should tend to make advertising-to-sales ratios increasing in revenues: many medical journals reach a fairly wide audience, so the cost per potential patient of journal advertising is probably decreasing in the size of the pool of potential patients. The competition effect should go in the other direction for the same reason as it did with detail advertising. The fact that the two effects go in opposite directions makes this a less than ideal application for our methodology, although we would expect that the direct effect would outweigh the competition effect and produce a monotone increasing pattern absent entry-deterrence motivations.

The prediction of the strategic-entry-deterrence model is as in the detail advertising case: firms should reduce advertising in intermediate-sized markets to reduce the attractiveness of the market to potential entrants.

The second column of Table 6 reports estimates from a regression of $Journal3/Revenue3$ on $\log(Revenue3)$, $(\log(Revenue3) - \bar{R})^2$ and $Specialist$. The positive coefficient on $\log(Revenue3)$ suggests that advertising-to-sales ratios are higher in larger markets, but this coefficient and all the others are not significant. The R^2 is again low.

The quintile means reported in the second row of Table 7 indicate that journal advertising is also lowest in the second quintile. This time, the Hall-Heckman test rejects monotonicity at the 5% level.

We would summarize these results by saying that they suggest that there is evidence of nonmonotonicity in the relationship between journal advertising and market size. The form of the nonmonotonicity is consistent with what one would expect from an entry-deterrence model. We should note, however, that one could also tell a nonstrategic story: it could be that the competition effect dominates in small markets whereas economies of scale drive the cross-sectional pattern in large markets.

6.1.3 Presentation proliferation

The final incumbent action we examine in this section is presentation proliferation. As noted above, the entry-deterrence motive gives incumbents in intermediate-sized markets an incentive to increase the number of presentations in which the drug is sold. This makes it more costly for entrants to match the incumbents' full product line (or leaves entrants with a lower market share if they enter with a limited product line). This would be reflected in a reduction in our Herfindahl-style measure, *PresHerf3*, in intermediate-sized markets.

This again is a not an ideal application because there are effects going in both directions. There should be a strong negative "direct effect" because the fixed costs of developing a new presentation can be spread over a larger quantity of sales in large markets. The "competition effect" likely goes in the opposite direction because the incumbent has lower sales in a duopoly. We would anticipate, however, that the latter is likely to be smaller, so we are comfortable applying our tests assuming that a monotone decreasing relationship would be most likely absent entry-deterrence motives.

The third column of Table 6 reports coefficient estimates from a regression of *PresHerf3* on $\log(\text{Revenue3})$, $(\log(\text{Revenue3}) - \bar{R})^2$, and two control variables. There is clearly a strong relationship between presentation proliferation and revenues: the coefficient on $\log(\text{Revenue3})$ is negative and highly significant. The quadratic term yields no evidence of nonmonotonicity. The control variables, *Psych* and *Topical*, are also highly significant.³⁵

³⁵Psychoactive drugs tend to be offered in a much wider range of dosages than other drugs and topical medications tend to have a large number of presentations both because of variation in the dosage and

Table 7 similarly gives no indication of a nonmonotone pattern. The quintile means are monotonically decreasing. The p-values for the two monotonicity tests are 0.34 and 0.19.

We would summarize these results by saying that they suggest that there is a strong direct effect of market size on presentation proliferation. Our approach will have less power to detect strategic entry deterrence when direct effects are larger. The cross-section data on presentation proliferation appear to be a case where our approach will have limited power.

6.2 Changes in behavior as expiration approaches

In this section we exploit an additional feature the pharmaceutical environment—entry is prohibited until a known point in time—to construct additional tests for strategic entry deterrence. We find additional evidence of nonmonotonicities.

One can think of the analysis in the previous subsection as examining a cross-section produced by a data generating process of the form

$$A_{it} = A_{ND}^*(z_i) + \gamma_t(A_{ED}^*(z_i) - A_{ND}^*(z_i)) + \eta_i + \epsilon_{it},$$

where we have broken the equilibrium behavior $A^*(z_i)$ into two components: the behavior that would have been optimal if firms ignored the strategic entry-deterrence motive, $A_{ND}^*(z_i)$, and the pure strategic term, $A_{ED}^*(z_i) - A_{ND}^*(z_i)$, multiplied by a coefficient γ_t , which would take on the value of one if firms fully recognized and reacted to the strategic entry-deterrence motive, and zero in the other extreme where firms do not recognize this incentive or choose not to respond to it. Writing the equation this way highlights two reasons why it might be difficult to find evidence of strategic entry deterrence in a pure cross-section dataset. First, if $A_{ND}^*(z_i)$ is steeply sloped, then this may overwhelm the nonmonotonicity of the strategic term $\gamma_t(A_{ED}^*(z_i) - A_{ND}^*(z_i))$ and leave us with no nonmonotonicity to detect. Second, even if this is not the case, the error terms due to drug-specific heterogeneity and noise, η_i and ϵ_{it} , may be large enough to make it hard to provide significant evidence from small sample sizes.

Suppose that one also had the opportunity to observe each market at two points in time: a time t when the incumbent invested as above and a second point in time t' when

because they may be offered as creams, liquids, gels, etc. in different-sized tubes.

the entry-deterrence motive was weaker (or absent) and the incumbent chose

$$A_{it'} = A_{ND}^*(z_i) + \gamma_{t'}(A_{ED}^*(z_i) - A_{ND}^*(z_i)) + \eta_i + \epsilon_{it'},$$

with $\gamma_{t'} < \gamma_t$. The difference between the two observations would then be given by

$$A_{it} - A_{it'} = (\gamma_t - \gamma_{t'})(A_{ED}^*(z_i) - A_{ND}^*(z_i)) + \epsilon_{it} - \epsilon_{it'}.$$

Detecting nonmonotone pattern in this equation could be much easier: the potentially steeply sloped nonstrategic term has been eliminated, and the error variance may be smaller.

In our pharmaceutical application, generic entry is prohibited until a known date. Further in advance of this date, the entry deterrence motive should be weaker, *e.g.* if the rate at which advertising goodwill decays is such that advertising today will have only a trivial impact on demand at the patent-expiration date, then the incentive to distort advertising will be very weak. Of course, if one looks much further from the patent expiration date it becomes less plausible that the nonstrategic optimum, $A_{ND}^*(z_i)$, and the drug-specific heterogeneity η_i , are really the same at the two points in time, *e.g.* firms advertise much more when launching a new drug than in subsequent years. To balance these two considerations, we examine here the difference between firm behavior in the year immediately prior to patent expiration and firm behavior in the preceding two years.³⁶

In the subsections that follow, we examine whether changes in each of the potential “strategic investments” are nonmontone in market size. For these tests, we drop all observations for which there is no change in incumbent behavior, *e.g.* drugs which do zero advertising in both years and drugs that are only ever offered in a single presentation. Our primary dependent variable is an indicator variable for the behavior having increased or decreased. This makes the comparisons as simple as possible and eliminates problems due to outliers and heteroskedasticity.

6.2.1 Detail advertising

Recall that a strategic entry-deterrence model predicts that firms in intermediate-sized markets would reduce their detail advertising to make their market less attractive to potential

³⁶In the case of the advertising variables, which are available to us at monthly frequency, we do this by comparing the twelve months prior to patent expiration with the preceding twenty-four months. Our presentation and pricing data are at an annual frequency, and we compare the calendar year prior to expiration with the two previous calendar years.

entrants. We did not see evidence of such a pattern in our cross-section analysis of detail advertising, though.

For a first look at the question of which firms are increasing and decreasing detail advertising prior to patent expiration, the first row of Table 8 reports the fraction of each drugs in each revenue quintile that increased their detail advertising in the year prior to patent expiration. (The row below this lists the number of drugs remaining in each quintile after we dropped drugs that did no detail advertising at all in the thirty-six months prior to patent expiration.) Note that a pattern roughly consistent with the entry-deterrence model is again visible: the majority of drugs in the smallest- and largest-revenue quintile are increasing detailing as patent expiration approaches; while most drugs in the second- and third quintiles (which are the ones where entry is most uncertain) are decreasing detail advertising. Although the sample sizes behind some of these numbers are small (and one could have worried that the stark numbers in the table were sensitive to the arbitrary bin cutoffs), we are able to strongly reject monotonicity with the nonparametric Ellison-Ellison style test.

We conclude that when we examine detail advertising changes prior to patent expiration instead of levels, there is evidence of strategic entry-deterrence: firms in intermediate-sized markets are reducing detail advertising prior to patent expiration.

6.2.2 Journal advertising

The strategic use of journal advertising is similar: firms in intermediate-sized markets would reduce journal advertising to deter entry. The second row of Table 8 indicates that reducing journal advertising is most common among firms in the third revenue-quintile. Neither of the nonparametric tests finds the departure from monotonicity to be significant at reasonable levels, though.

We conclude with a mixed message. Our cross-section analysis found evidence that journal advertising was being altered for entry-deterrence reasons. Looking at changes in journal advertising as patent expiration approaches, though, does not provide additional evidence of this strategic behavior.

6.2.3 Presentation proliferation

The potential strategic use of presentation proliferation is that firms in intermediate-sized markets could try to deter entry by introducing more presentations. This would mean that our *PresHerf* measure of presentation dispersion would be lower in these markets.

The quintile means reported in the third main row of Table 8 suggest the possibility of nonmonotonicity, but the evidence is not conclusive. The Hall-Heckman test indicates a departure from monotonicity significant at about the 8% level.

We conclude that looking at how firms change their behavior in the year prior to patent expiration yields some additional weaker evidence that firms might be influenced by a strategic entry-deterrence motive, this time in the context of presentation proliferation.

6.2.4 Pricing

Pricing is an area in which looking at changes in behavior has the largest incremental benefit. We did not attempt to discuss strategic pricing at all in our cross-section analysis because it is hard to normalize prices in any way that makes comparisons across drugs meaningful.³⁷ Looking at whether firms are increasing or decreasing prices as patent expiration approaches, in contrast, is both simple and sensible.

What patterns might one see if pricing decisions are influenced by an entry deterrence motive? As mentioned above, in the simplest limit-pricing models, firms set low prices to deter entry, and this is effective because firms are somehow committed to these prices. This story could make sense in the pharmaceutical industry, with government or public scrutiny being the external force that makes raising prices costly. With regard to the various signalling and signal-jamming theories of strategic pricing, our discussions with industry sources suggest that generic firms are well-informed about both prices and revenues (and indeed have all the same data we have). Generic firms are less likely to be well informed about price elasticities. Hence, it might be plausible to imagine that firms could choose prices that are too low from the perspective of static profit maximization in order to convince

³⁷What one would want for a study of strategic pricing is to look at each drug's price relative to the level that would be optimal absent entry-deterrence motives, but any normalization of this kind would require much more information than is available to us, *e.g.* one would want to estimate own-price elasticities for each drug.

generic entrants that elasticities are such that it will be profitable for them to continue to charge low prices after generic entry. In summary, the most plausible applications of strategic pricing models suggest that firms in intermediate-sized markets might distort prices downward to deter entry.³⁸

The fourth row of Table 8 reports the frequency with which firms in each quintile are raising prices charged to drugstores in the year before patent expiration.³⁹ The quintiles suggest some nonmonotonicity, but the formal nonparametric tests do not find it to be significant.

The fifth row repeats this exercise for hospital prices. Here, the general pattern is that price increases are more common in the higher-revenue quintiles and the monotonicity tests indicate that there is no significant evidence of nonmonotonicity.

7 Conclusion

The expiration of a pharmaceutical patent and the subsequent opening of a drug market to potential entrants is a momentous event for pharmaceutical firms. In this paper we have examined how a number of firms have set prices, chosen advertising levels, and adjusted their presentation-level product mix at this time. In some cases, we have found evidence of nonmonotonic patterns that suggest that incumbents' actions may be motivated in part by a desire to deter generic entry. We found some evidence of nonmonotonic behavior in journal advertising in the cross-section and of detail advertising and presentation proliferation in changes before patent expiration. One possible interpretation of this pattern of results is that journal advertising is considered to be a longer-lived investment than the other two strategic tools, so that any strategic use of journal advertising could precede patent

³⁸Incumbents distorting their prices down in advance of entry might also provide an additional explanation for the much talked about observation that incumbents sometimes raise prices following generic entry. See Masson and Steiner (1985), Hurwitz and Caves (1988), Caves, Whinston and Hurwitz (1991), Grabowski and Vernon (1992) and Griliches and Cockburn (1994). One situation in which the opposite distortion in prices might be expected is when the incumbent also sells another product in the therapeutic category that has a greater remaining patent life. In such a situation, a strategy for dealing with generic entry which has been mentioned to us is to try to induce consumers of the product with the expiring patent to switch to the other product. One way to do this is to raise the price of the older product.

³⁹To be precise, the table reports the fraction of drugs for which the real price was higher in the year immediately prior to patent expiration than it was, on average, in the two previous years. Note that not all drugs in our sample were sold in both hospitals and drugstores.

expiration by a relatively long time.

From a practical perspective, it should be pointed out that the entry-detering behavior we identify is not in blockbuster drugs—it occurs for more obscure drugs with relatively low revenues. This is a result of our empirical strategy, and should not be interpreted as a suggestion that strategic behavior is more common in these incumbents than in incumbents with blockbuster drugs. On the contrary, the fact that these incumbents seem to have figured out the potential gains from altering future competitive conditions on these unimportant drugs would lead one to assume that firms are also strategically sophisticated with regard to their more important products.

More generally, this is a paper about the testing of strategic entry-deterrence theories. Strategic investment models have become widespread in industrial organization, strategic management, and other fields over the last two decades. Empirical analyses of such models should be useful for diverse reasons: from a behavioral perspective one could wonder whether firms have figured out the sometimes subtle effects; and regulators may be interested in whether firms are actively trying to deter entry. Direct tests of strategic intent are made difficult by the need to precisely estimate long run elasticities and to consider the value of investments in alternate states of the world, however, and the empirical literature to-date is limited. Our approach does not require extensive data. We hope that it may thereby enable future work in this area.⁴⁰

We hope also that our paper may more generally spur future work on monotonicity tests as a tool for applied work. Predictions about monotonicity might be used to distinguish between theories in other sorts of models as well. The robustness of the approach to common econometric difficulties could be a significant advantage.

⁴⁰See Dafny (2005) for an interesting application to hospital markets.

Appendix

A. Proof of Proposition 1

The first order condition for $A_{ND}^*(z)$ is

$$\frac{\partial c}{\partial A}(A_{ND}^*(z), z) = F(\pi_2^{d^*}(A_{ND}^*(z), z)) \frac{\partial \pi_1^{d^*}}{\partial A}(A_{ND}^*(z), z) + (1 - F(\pi_2^{d^*}(A_{ND}^*(z), z))) \frac{\partial \pi_1^{m^*}}{\partial A}(A_{ND}^*(z), z).$$

Differentiating with respect to z gives

$$\begin{aligned} \frac{\partial^2 c}{\partial A^2} \frac{dA^*}{dz} + \frac{\partial^2 c}{\partial z \partial A} &= F(\pi_2^*) \left(\frac{\partial^2 \pi_1^{d^*}}{\partial z \partial A} + \frac{\partial^2 \pi_1^{d^*}}{\partial A^2} \frac{dA^*}{dz} \right) + (1 - F(\pi_2^*)) \left(\frac{\partial^2 \pi_1^{m^*}}{\partial z \partial A} + \frac{\partial^2 \pi_1^{m^*}}{\partial A^2} \frac{dA^*}{dz} \right) \\ &\quad + f(\pi_2^{d^*}(A_{ND}^*(z), z)) \frac{d\pi_2^{d^*}}{dz} \left(\frac{\partial \pi_1^{d^*}}{\partial A} - \frac{\partial \pi_1^{m^*}}{\partial A} \right), \end{aligned}$$

where we have written $\frac{d\pi_2^{d^*}}{dz}$ for the total derivative of $\pi_2^{d^*}(A_{ND}^*(z), z)$ with respect to z , π_2^* for $\pi_2^{d^*}(A_{ND}^*(z), z)$, and where all derivatives are evaluated at $(A_{ND}^*(z), z)$.

Solving for $\frac{dA^*}{dz}$ gives

$$\frac{dA^*}{dz} = \frac{F(\pi_2^*) \frac{\partial^2 \pi_1^{d^*}}{\partial z \partial A} + (1 - F(\pi_2^*)) \frac{\partial^2 \pi_1^{m^*}}{\partial z \partial A} - \frac{\partial^2 c}{\partial z \partial A} + f(\pi_2^*) \frac{d\pi_2^{d^*}}{dz} \left(\frac{\partial \pi_1^{d^*}}{\partial A} - \frac{\partial \pi_1^{m^*}}{\partial A} \right)}{\frac{\partial^2 c}{\partial A^2} - F(\pi_2^*) \frac{\partial^2 \pi_1^{d^*}}{\partial A^2} - (1 - F(\pi_2^*)) \frac{\partial^2 \pi_1^{m^*}}{\partial A^2}},$$

where again all derivatives are evaluated at $(A_{ND}^*(z), z)$.

The denominator of this expression is always positive. Given the assumption that $\frac{d\pi_2^{d^*}}{dz} > 0$, the numerator is a sum of the direct effect and the product of the competition effect and something that is nonnegative. Hence, $A_{ND}^*(z)$ will be monotone increasing if the two effects are positive and monotone decreasing if they are both negative.

QED.

B. Monte Carlo Study of Monotonicity Tests

Table 1 reports results of simulations designed to assess the size and power of the tests of monotonicity we use.

The table reports the frequency with which monotonicity was rejected when we constructed simulated datasets using various data generating processes and then tested for monotonicity via the same procedures we apply to our real data. To compute each entry in the table we constructed 1000 simulated datasets, ran our bootstrap procedure to find estimated 5% critical values, and then compared the value of the test statistic on the simulated dataset to the estimated critical value.⁴¹ The simulated datasets had 100 data points with uniform draws on $[0, 1]$. The first column reports rejection rates (based on 5% critical

⁴¹We used 1000 bootstrap repetitions to construct the estimated critical values. The procedure for doing these was to fit an isotone regression to the simulated dataset and draw errors from the difference between the actual and fitted values.

values) for the Hall-Heckman style test with smoothing parameter $m = 15$. The second column reports rejection rates for the Ellison-Ellison style test with window width $w = 0.2$.

The first three rows of the table examine the size of the test statistics under three different monotone data generating processes. The first, labeled M1, is $A_i = \epsilon_i$, with ϵ_i a standard normal random variable. The second, M2, is $A_i = x(2-x) + \epsilon_i$. In the third, M3, A_i is a 0/1 variable generated from the linear probability model $\text{Prob}\{A_i = 1|z_i\} = 0.25 + 0.5z_i$. The rejection rates are usually around 5% as they should be.

The fourth and fifth rows examine the power of the tests when applied to nonmonotone data generating processes. The process labelled NM1 consisted of setting $A_i = 10z_i(1.4 - z_i) + \epsilon_i$, with ϵ_i a standard normal random variable. Note that this function is increasing on $[0, 0.7)$ and decreasing on $(0.7, 1]$. The value at the right endpoint is 0.9 less than the value at the peak, which is slightly less than one standard deviation of the error distribution. In the process labelled NM2, A_i is a 0/1 variable with $\text{Prob}\{A_i = 1|z_i\} = 0.25 + 2z_i(1 - z_i)$. The rejection rates in these simulations range from 26.8% to 47.2%. Hence, one can think of processes NM1 and NM2 as exemplifying the magnitude of the departure from monotonicity that must be present in the data to be detected by our tests.

Table 1: Simulation study of size and power of monotonicity tests

Data generating process	Rejections with 5% critical values	
	H-H based test	E-E based test
M1	0.056	0.056
M2	0.062	0.048
M3	0.038	0.018
NM1	0.356	0.398
NM2	0.268	0.472

The table presents rejection rates for the Hall-Heckman and Ellison-Ellison style monotonicity tests when applied to five data generating processes with 5% critical values obtained from a bootstrap procedure.

C. List of Drugs Used in the Study

Generic name	Brand names	Expiration	Revenue3	Entry3Yr
albuterol	proventil	1989	172,952	1
	ventolin			
amiodarone hydrochloride	cordarone	1990	11,283	0
amoxapine	asendin	1989	18,306	1
atenolol	tenormin	1991	301,311	1
	tenoretic			
auranofin	ridaura	1992	9,766	0
baclofen	lioresal	1986	12,033	1
betamethasone	celestone	1986	8,226	0
bretylum tosylate	bretylol	1986	10,418	1
bromocriptine mesylate	parlodel	1990	54,031	0
carbidopa	sinemet	1991	91,883	1
carboprost tromethamine	hemabate	1990	189	0
chlorpheniramine maleate	ornade	1986	15,303	1
chlorthalidone	combipres	1986	15,988	1
cinoxacin	cinobac	1989	5,680	1
clonidine	catapres	1986	70,045	1
clorazepate dipotassium	tranxene	1987	87,533	1
clotrimazole	gyne-lotrimin	1989	55,283	0
	mycelex			
	lotrimin			
colestipol hydrochloride	colestid	1989	7,089	0
cromolyn sodium	nasalcrom	1989	49,640	1
	intal			
cyclobenzaprine hydrochlor	flexeril	1986	40,630	1
cytarabine	cytosar	1986	8,140	1
deferoxamine mesylate	desferal	1986	3,366	0
desipramine hydrochloride	norpramin	1986	19,439	1
desmopressin acetate	ddavp	1987	6,112	0
dimethyl sulfoxide	rimso50	1987	296	0
dipivefrin hydrochloride	propine	1991	23,353	1
doxepin hydrochloride	adapin	1986	55,059	1
	sinequan			
dronabinol	marinol	1990	1,454	0
enflurane	ethrane	1987	19,337	1
fenoprofen calcium	nalfon	1988	49,538	1

cont'd

Generic name	Brand names	Expiration	Revenue3	Entry3Yr
fluocinonide	lidex	1988	24,012	1
fluorometholone	fluor-op	1989	93	0
flurandrenolide	cordran	1989	4,527	0
guanfacine hydrochloride	tenex	1991	23,530	0
halazepam	paxipam	1986	1,621	0
haloperidol	haldol	1986	72,705	1
ipratropium bromide	atrovent	1991	37,356	0
ketoprofen	orudis	1991	60,313	1
loperamide hydrochloride	imodium	1990	28,278	1
loxapine hydrochloride	loxitane	1987	11,567	1
mazindol	mazanor	1990	2,001	0
	sanorex			
mebendazole	vermox	1989	6,154	0
metaproterenol sulfate	alupent	1986	39,260	1
miconazole	monistat	1991	107,102	1
molindone hydrochloride	moban	1987	2,725	0
nalidixic acid	neggram	1988	4,501	1
naloxone hydrochloride	narcan	1986	15,262	1
naltrexone	trexan	1989	723	0
norgestrel	ovrette	1991	637	0
pancuronium bromide	pavulon	1988	18,801	1
piroxicam	feldene	1992	216,998	1
prazosin hydrochloride	minipress	1989	67,923	1
procarbazine hydrochloride	matulane	1987	565	0
pyrantel	antiminth	1989	597	0
stanozolol	winstrol	1989	541	0
sulfasalazine	azulfidine	1988	9181	1
sulindac	clinoril	1990	164,545	1
timolol maleate	blocadren	1989	114,148	1
	timoptic			
tolmetin sodium	tolectin	1990	48,654	1
tretinoin	retin a	1990	61,167	0
trilostane	modrastane	1989	24	0
verapamil hydrochloride	isoptin	1986	56,494	1
	calan			

References

- Aghion Philippe, and Bolton, Patrick (1987): "Entry Prevention Through Contracts with Customers," *American Economic Review*, 77, 388-401.
- Aguirregabiria, V. and Mira, P. (2007): "Sequential Estimation of Discrete Games," forthcoming *Econometrica*.
- Bae, J. P. (1997): "Drug Patent Expirations and the Speed of Generic Entry," *Health Services Research* 32, 87-101.
- Bajari, P., Benkard, C.L., and Levin, J. (2006): "Estimating Dynamic Models of Imperfect Competition," mimeo.
- Berndt, Ernst, Bui, Linda, Reiley, David, and Urban, Glen (1995): "Information, Marketing, and Pricing in the U.S. Anti-Ulcer Drug Market," *American Economic Association Papers and Proceedings*, 85, 100-105
- Bronnenberg, Bart, Dhar, Sanjay, and Dube, Jean-Pierre (2005): "Endogenous Sunk Costs and the Geographic Distribution of Brand Shares in CPG Industries, mimeo, University of Chicago.
- Bulow, Jeremy, Geanakoplos, John, and Klemperer, Paul (1985): "Multimarket Oligopoly: Strategic Substitutes and Complements," *Journal of Political Economy*, 93, 488-511.
- Caves, Richard, Whinston, Michael, and Hurwitz, M. (1991): "Patent Expiration, Entry, and Competition in the U.S. Pharmaceutical Industry," *Brookings Papers on Economic Activity*, 1-66.
- Chevalier, Judith (1995a): "Capital Structure and Product Market Competition: Empirical Evidence from the Supermarket Industry," *American Economic Review*, 85, 415-435.
- Chevalier, Judith (1995b): "Do LBO Supermarkets Charge More? An Empirical Analysis of the Effects of LBOs on Supermarket Pricing," *Journal of Finance* 50, 1095 - 1112.
- Cooper, David J., Susan Garvin and John Kagel (1997): "Signalling and Adaptive Learning in an Entry Limit Pricing Game," *RAND Journal of Economics*, 28, 662-683.
- Cooper, T.E. (1986): "Most-Favored Customer Pricing and Tacit Collusion," *RAND Journal of Economics*, 17, 377-388.
- Dafny, Leemore (2005): "Games Hospitals Play: Entry Deterrence in Hospital Procedure Markets," *Journal of Economics and Management Strategy*, 14, 513-542.
- Dixit, Avinash (1980): "The Role of Investment in Entry Deterrence," *Economic Journal*, 90, 95-106.
- Ellison, Glenn, and Ellison, Sara Fisher (2000): "A Simple Framework for Nonparametric Specification Testing," *Journal of Econometrics*, 96, 1-23.

- Ellison, Sara Fisher, Cockburn, Iain, Griliches, Zvi, and Hausman, Jerry (1997): "Characteristics of Demand for Pharmaceutical Products: An Examination of Four Cephalosporins," *RAND Journal of Economics*, 28, 426-446.
- Ellison, Sara Fisher, and Snyder, Christopher (2001): "Countervailing Power in Wholesale Pharmaceuticals," MIT Economics Department Working Paper 01-27.
- Ellison, Sara Fisher, and Wolfram, Catherine (2006): "Coordinating on Lower Prices: Pharmaceutical Pricing Under Political Pressure," *RAND Journal of Economics*, 37, 324-340.
- Fudenberg, Drew, and Tirole, Jean (1983a): "Capital as a Commitment: Strategic Investment to Deter Mobility," *Journal of Economic Theory*, 31, 227-250.
- Fudenberg, Drew, and Tirole, Jean (1983b): "Learning by Doing and Market Performance," *Bell Journal of Economics*, 14, 522-530.
- Fudenberg, Drew, and Tirole, Jean (1984): "The Fat Cat Effect, the Puppy Dog Ploy, and the Lean-and-Hungry Look," *American Economic Review*, 74, 361-366.
- Fudenberg, Drew, and Tirole, Jean (1986): "A 'Signal Jamming' Theory of Predation," *Rand Journal of Economics*, 17, 366-376.
- Gaskins, Darius W., Jr. (1971): "Dynamic Limit Pricing: Optimal Pricing Under Threat of Entry," *Journal of Economic Theory* 3, 306-322.
- Gelman, J.R., and Salop, S.C. (1983): "Judo Economics: Capacity Limitation and Coupon Competition," *Bell Journal of Economics*, 14, 315-325.
- Ghosal, Subhashis, Sen, Arusharka, and van der Vaart, W. (2000): "Testing Monotonicity in Regression," *Annals of Statistics*, 28, 1054-1082.
- Gijbels, Irene, Hall, Peter, Jones, M.C., and Koch, Inge (2000): "Tests for Monotonicity of a Regression Mean with Guaranteed Level," *Biometrika*, 87, 663-673.
- Goolsbee, Austan and Syverson, Chad (2004): "How Do Incumbents Respond to the Threat of Entry? Evidence from the Major Airlines?," mimeo, University of Chicago.
- Grabowski, H., and Vernon, J. (1992): "Brand Loyalty, Entry, and Price Competition in Pharmaceuticals after the 1984 Drug Act," *Journal of Law and Economics* 35, 331-350.
- Grabowski, H., and Vernon, J. (1996): "Longer Patents for Increased Generic Competition in the U.S.: the Waxman-Hatch Act after One Decade," *PharmacoEconomics* 10, 110-123.
- Griliches, Zvi, and Cockburn, Iain (1994): "Generics and New Goods in Pharmaceutical Price Indexes," *American Economic Review*, 84, 1213-1232.
- Hall, Peter, and Heckman, Nancy (2000): "Testing for Monotonicity of a Regression Mean by Calibrating for Linear Functions," *Annals of Statistics*, 28, 20-39.
- Harrington, Joseph (1986): "Limit Pricing When the Potential Entrant is Uncertain of Its Cost Function," *Econometrica* 54, 429-437.

- Hurwitz, M. and Caves, Richard (1988): "Persuasion or Information? Promotion and the Shares of Brand Name and Generic Pharmaceuticals," *Journal of Law and Economics*, 31, 299-320.
- Judd, Kenneth (1985): "Credible Spatial Preemption," *RAND Journal of Economics*, 16, 153-166.
- Kadiyali, Vrinda (1996): "Entry, Its Deterrence and Its Accommodation: A Study of the U.S. Photographic Film Industry," *Rand Journal of Economics* 27, 452-478.
- Klemperer, Paul (1987): "The Competitiveness of Markets with Switching Costs," *Rand Journal of Economics* 18, 138-150.
- Leffler, K. (1981): "Persuasion or Information? The Economics of Prescription Drug Advertising," *Journal of Law and Economics* 24, 45-74.
- Lieberman, Marvin (1987): "Post Entry Investment and Market Structure in the Chemical Processing Industry," *RAND Journal of Economics*, 18, 533-549.
- Masson, A. and Steiner, R. (1985): *Generic Substitution and Prescription Drug Prices: Economic Effects of State Drug Product Selection Laws*. U.S. Federal Trade Commission.
- Milgrom, Paul (1981): "Good News and Bad News: Representation Theorems and Applications," *The Bell Journal of Economics* 12, 380-391.
- Milgrom, Paul and Roberts, J. (1982): "Limit Pricing and Entry Under Incomplete Information: An Equilibrium Analysis," *Econometrica* 50, 443-460.
- Rizzo, J. (1999): "Advertising and Competition in the Ethical Pharmaceutical Industry: The Case of Hypertensive Drugs," *The Journal of Law and Economics* 42, 89-116.
- Rosenthal, Meredith, Berndt, Ernst, Donohue, Julie, Frank, Richard, and Epstein, Arnold (2002): "Promotion of Prescription Drugs to Consumers," *New England Journal of Medicine*, 346, 498-505.
- Schmalensee, Richard (1978): "Entry Deterrence in the Ready-to-Eat Breakfast Cereal Industry," *Bell Journal of Economics*, 9, 305-27.
- Schmalensee, Richard (1981): "Economies of Scale and Barriers to Entry," *Journal of Political Economy*, 89, 1228-38.
- Schwartzman, D. (1976): *Innovation in the Pharmaceutical Industry*. The Johns Hopkins University Press.
- Schweitzer, S. (1997): *Pharmaceutical Economics and Policy*, Oxford University Press.
- Scott Morton, Fiona (1999): "Entry Decisions in the Generic Drug Industry," *RAND Journal of Economics* 30, 421-440.
- Scott Morton, Fiona (2000): "Barriers to Entry, Brand Advertising and Generic Entry in the U.S. Pharmaceutical Industry," *International Journal of Industrial Organization* 18,

1085-1104.

Smiley, R. (1988): "Empirical Evidence on Strategic Entry Deterrence," *International Journal of Industrial Organization*, 6, 167-180.

Sorensen, Alan (2000): "Equilibrium Price Dispersion in Retail Markets for Prescription Drugs," *Journal of Political Economy* 108, 833-850.

Temin, Peter (1980): *Taking Your Medicine: Drug Regulation in the United States*. Harvard University Press.

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

U.S. Congress, Office of Technology Assessment (1993): *Pharmaceutical R&D: Costs, Risks, and Rewards*. U.S. Government Printing Office.

Walker, H. (1971): *Market Power and Price Levels in the Ethical Drug Industry*. Indiana University Press.

Model of Advertising with Spillovers

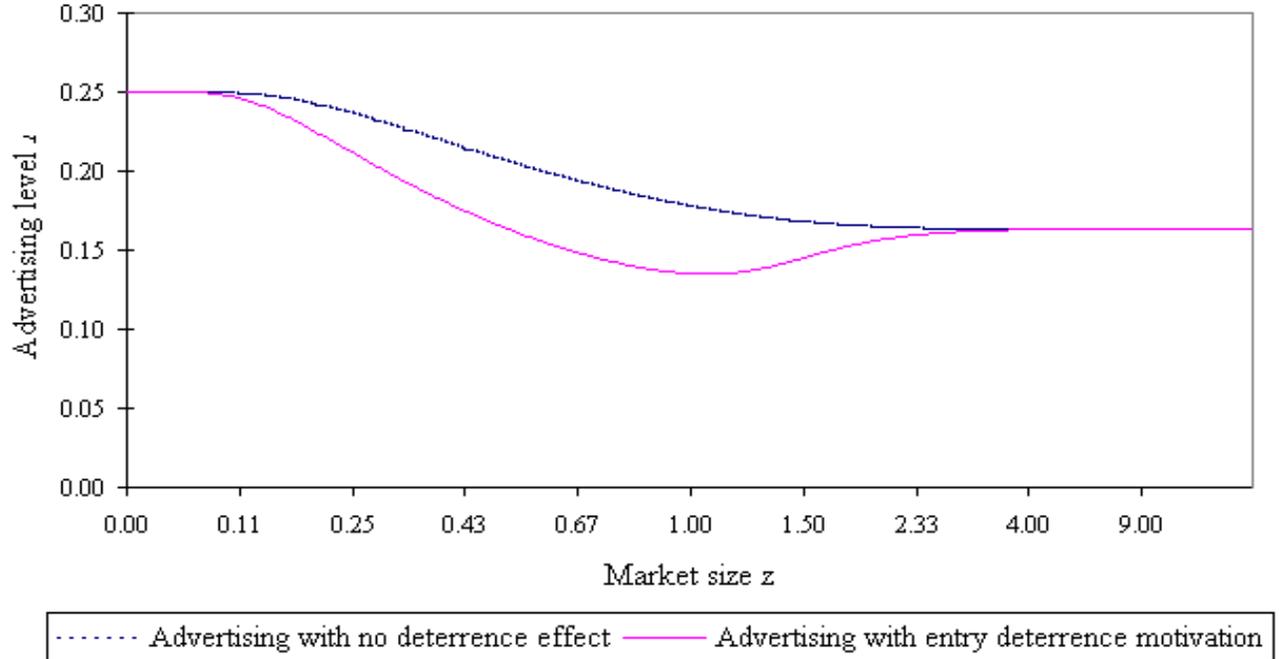


Figure 2: Equilibrium advertising levels in the model of advertising with spillovers

The figure graphs the equilibrium advertising intensity in the model of section 2.3 where advertising raises consumers valuations both for the branded drug and for a generic substitute. The distribution of entry costs is assumed to be lognormal with mean 0.0025 and standard deviation 0.0015. The dotted line is the equilibrium advertising level when advertising is not observed until after firm 2's entry decision is made (and hence there is no entry deterrence motive.) The solid line is the equilibrium advertising level when advertising is observed in advance of the potential entry.

Table 2: Variable names

Variable Name	Variable Description
<i>Entry3Yr</i>	1 if entry within 3 years of patent expiration
<i>EntryProb</i>	Predicted entry probability
<i>Chronic</i>	0 if for acute illness; 1 if for chronic illness
<i>HospFrac</i>	Hospital fraction of revenue (for year prior to patent expiration)
<i>Revenue3</i>	Average annual revenue for 3 years prior to patent expiration (000's constant dollars)
<i>TherSubs</i>	Number of other drugs in the therapeutic class
<i>Detail</i>	Monthly detailing advertising (000's of minutes)
<i>Journal</i>	Monthly journal advertising expenditures (000's of constant dollars)
<i>Detail3</i>	Average annual detailing in 3 years before patent expiration
<i>Journal3</i>	Average annual journal advertising in 3 years before patent expiration
<i>PresHerf</i>	<i>HospFrac</i> -weighted average of drugstore and hospital presentation Herfindahls
<i>PresHerf3</i>	Average of <i>PresHerf</i> in the 3 years before patent expiration
<i>HPrice</i>	Hospital price (in constant dollars)
<i>DPrice</i>	Drugstore price (in constant dollars)
<i>Specialist</i>	Index for how often drugs in therapeutic class are prescribed by specialist
<i>Psych</i>	1 if drug is psychoactive
<i>Topical</i>	1 if drug is applied topically

The table describes the variables used in the analysis.

Table 3: Summary statistics

Variable	Number of Observations	Mean	Standard Deviation
<i>Entry3Yr</i>	63	0.59	0.50
<i>Revenue3</i>	63	39,355	55,754
$\log(\textit{Revenue3})$	63	9.40	2.00
<i>HospFrac</i>	63	0.21	0.30
<i>Chronic</i>	63	0.63	0.42
<i>TherSubs</i>	63	8.48	6.04
<i>Detail3/Revenue3</i>	69	0.005	0.008
<i>Journal3/Revenue3</i>	70	0.014	0.022
<i>PresHerf3</i>	70	0.54	0.29
$DPrice_t/DPrice_{t-1}$	245	1.019	0.067
$HPrice_t/HPrice_{t-1}$	233	1.010	0.129

The table presents summary statistics for some of the variables used in our analysis.

Table 4: Summary statistics by revenue quintile

Variable	Mean within revenue quintile				
	Q 1	Q 2	Q 3	Q 4	Q 5
<i>Revenue3</i>	882	7,572	22,161	52,336	127,359
<i>Entry3Yr</i>	0.00	0.43	0.86	0.80	0.92
Number of Obs.	13	14	14	10	12

The table reports the mean annual revenues in the three years prior to patent expiration and the fraction of drugs experiencing entry for drugs in each revenue quintile.

Table 5: Entry versus pre-expiration revenues

Dependent variable for probit is <i>Entry3Yr</i>		
Variable		
$\log(\textit{Revenue3})$	0.70 (0.17)	0.76 (0.20)
<i>HospFrac</i>		1.01 (0.78)
<i>Chronic</i>		0.60 (0.54)
$\log(\textit{TherSubs})$		0.01 (0.29)
Constant	-6.39 (1.94)	-7.60 (1.94)
Number of Obs.	63	63
<i>PseudoR</i> ²	0.40	0.43

The table presents estimates of probit models. The dependent variable is a dummy for whether entry occurs within three years of patent expiration. The explanatory variables are average revenue in the three years prior to patent expiration, the fraction of sales which are through hospitals (as opposed to drugstores), a measure of whether the drug treats a chronic or acute condition, and the number of other drugs in the therapeutic class. The observations are 63 drug molecules which lost patent protection at some point between 1986 and 1992.

Table 6: Incumbent behavior versus market size: linear regressions

Independent Variables	Dependent variable:		
	$\frac{Detail3}{Revenue3}$	$\frac{Journal3}{Revenue3}$	$PresHerf3$
$\log(Revenue3)$	0.000 (0.001)	0.003 (0.002)	-0.069 (0.016)
$(\log(Revenue3) - \bar{R})^2$	-0.0001 (0.0002)	0.0000 (0.0005)	-0.003 (0.005)
<i>Specialist</i>	0.006 (0.009)	0.008 (0.026)	
<i>Psych</i>			-0.342 (0.075)
<i>Topical</i>			-0.388 (0.090)
<i>Constant</i>	-0.014 (0.014)	-0.014 (0.041)	0.990 (0.360)
Number of Obs.	69	70	70
R^2	0.04	0.06	0.52

The table reports coefficient estimates from linear regressions of three types of investment, two advertising-to-sales ratios and the Herfindahl index of presentations, on the average revenue in the three years prior to patent expiration, the square of this variable minus its mean, and appropriate controls. The unit of observation is branded drugs which lost patent protection between 1986 and 1992.

Table 7: Incumbent behavior versus market size: quintile means and monotonicity tests

Variable	Variable mean for drugs in revenue quintile					Monotonicity test p -value	
	Q 1	Q 2	Q 3	Q 4	Q 5	H-H Test	E-E Test
$Detail3/Revenue3$	0.0051	0.0012	0.0055	0.0084	0.0041	0.274	0.161
$Journal3/Revenue3$	0.011	0.005	0.011	0.024	0.018	0.053	0.197
$PresHerf3$	0.78	0.64	0.49	0.44	0.35	0.336	0.187

The table reports the means of three types of investment, two advertising measures and the Herfindahl index of presentations, by revenue quintiles. Drugs are classified into quintiles based on the mean of their revenue for the three years prior to patent expiration. The EE and HH test columns reports the p -values for two tests of non-monotonicity (Ellison and Ellison 2000, Hall and Heckman 2000).

Table 8: Changes in incumbent behavior as expiration approaches: quintile means and monotonicity tests

	Fraction increasing by quintile					Monotonicity test p-value	
	Q 1	Q 2	Q 3	Q 4	Q 5	H-H Test	E-E Test
<i>Detail3</i>	0.75 (4)	0.22 (9)	0.25 (12)	0.54 (13)	0.62 (13)	0.307	0.031
<i>Journal3</i>	0.50 (2)	0.43 (7)	0.17 (12)	0.29 (14)	0.31 (13)	0.321	0.696
<i>PresHerf</i>	0.33 (6)	0.42 (12)	0.38 (13)	0.50 (14)	0.62 (13)	0.083	0.217
<i>DPrice</i>	0.70 (10)	0.58 (12)	0.75 (12)	0.54 (13)	0.92 (13)	0.430	0.601
<i>HPrice</i>	0.50 (8)	0.50 (12)	0.54 (13)	0.77 (13)	0.73 (11)	0.573	0.854

This table reports the fraction of drugs in each revenue quintile for which the investment variable was higher in the year immediately prior to patent expiration than it was on average in the previous two years. The number of observations in each cell is in parentheses below the quintile means.