American Economic Journal: Microeconomics 3 (February 2011): 1–36
http://www.aeaweb.org/articles.php?doi=10.1257/mic.3.1.1

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

By Glenn Ellison and Sara Fisher Ellison*

This paper develops a new approach to testing for strategic entry deterrence and applies it to the behavior of pharmaceutical incumbents before patent expiration. It examines a cross section of markets, determining whether behavior is nonmonotonic in market size. Under some conditions, investment levels will be monotone in market size if firms do not invest to deter entry. Strategic investments to deter entry, however, may result in nonmonotonic investment because they are unnecessary in small markets, and impossible in large ones. Consistent with an entry-deterrence motivation is the finding that incumbents in medium-sized markets advertise less prior to patent expiration. (*JEL* D92, G31, L11, L21, L65)

The insight that firms may make "strategic investments" to alter future competitive conditions is one of the most fundamental ideas in industrial organization. Jean Tirole's (1988) chapter reviewing arguments about how excess capacity, capital structure, advertising, contractual practices, learning-by-doing, and other actions can be used to deter entry is easily the longest in the text.[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 US 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.

* G. Ellison: Department of Economics, Massachusetts Institute of Technology, 50 Memorial Drive, Cambridge, MA 02142 and National Bureau of Economic Research (e-mail: gellison@mit.edu); S. F. Ellison: Department of Economics, Massachusetts Institute of Technology, 50 Memorial Drive, Cambridge, MA 02142 (e-mail: 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, Alan Sorensen, and Hongkai Zhang 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.

† To comment on this article in the online discussion forum, or to view additional materials, visit the article page at http://www.aeaweb.org/articles.php?doi=10.1257/mic.3.1.1.

[1] See Drew Fudenberg and Tirole (1984), and Jeremy I. Bulow, John D. Geanakoplos, and Paul D. Klemperer (1985) for frameworks organizing the literature.

In Section I, we begin with a discussion of strategic entry-deterrence models. We modify the textbook model to assume that entry costs are random, and unknown to the incumbent when the investment is made. We note that the incentive to deter entry will often 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. This can lead to a nonmonotonic relationship between equilibrium investment levels and market size.

At a general level, one could describe our approach to testing whether investment patterns suggest an entry-deterrence motive as follows: first, identify a prediction of the theory that would differ depending on whether firms take entry-deterrence benefits into account, and then examine the competing predictions in the data. Section II develops a taxonomy of strategic investment models in which investments are classified on the basis of whether "direct" and "competition" effects are positive or negative. We show that investments will be monotonically related to market size *if* firms do not have entry-deterrence motives; and the nature of the investment is such that the direct and competition effects have the same sign. The important implication of the theoretical result is, that *if* the two effects would go in the same direction in a given application, then one can reject the null hypothesis of no entry deterrence by testing and rejecting the hypothesis that there is a monotonic relationship between firms' actions and market size in a cross section of markets. (The requirement that the two effects go in the same direction is restrictive: in many applications, one would expect the two effects to work in opposite directions.)

Section III contains a brief discussion of the econometric literature on testing whether relationships are monotone, 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 IV, we turn our attention to the pharmaceutical industry. In addition to its practical importance, the pharmaceutical industry has two features that make it a nice environment in which to study strategic entry deterrence: there are several tools that incumbents might use to deter entry; and we can construct a cross-section dataset containing markets of different sizes. We have data on four potential strategic investments: "detail advertising" is sending representatives to doctors' offices; "journal advertising" is the placement of advertisements in medical journals; "presentation proliferation" refers to selling a drug in many different forms; and pricing. We treat every branded drug with a patent that is about to expire as a market in which an incumbent is threatened by potential entry. We treat revenue received in the United States in the years immediately prior to patent expiration as a proxy for market size because it is a strong predictor of whether generic entry will occur. The lowest-revenue drugs in our sample are unlikely to ever face generic competition, whereas entry is a near certainty for high revenue drugs. Such heterogeneity is important for our approach.

Section V presents our cross-section analysis of the behavior of pharmaceutical incumbents. For each of the potential strategic investments, we first present descriptive evidence on what incumbents do as their patents are close to expiration. Then, we comment on whether the investment satisfies the conditions under which monotonicity tests may be a useful way to look for evidence of strategic behavior. Some

do, and some do not. Finally, we look for evidence of entry deterrence by formally testing whether the behavior is nonmonotonically related to pre-expiration revenues in the cross section, and informally discussing whether any apparent nonmonotonicities are of the form that would be expected in a strategic entry deterrence model.

Section VI exploits the panel nature of our dataset as another potential source of evidence of entry-deterrence motives. While our approach to testing for strategic entry deterrence only requires a single cross section, we note that 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. In theory, such an approach may make tests more powerful. In our application, however, we find fewer significant estimates here. We conclude that there are some patterns in the data that would be consistent with firms' reacting to entry-deterrence motives, but that the evidence is not very strong. 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.[2] Most closely related is Richard E. Caves, Michael D. Whinston, and Mark A. Hurwitz (1991), a descriptive study based on 30 drugs with patents expiring between 1976 and 1987. The authors look mostly at the average behavior of incumbents before and after expiration. Also, drugs in the study are separated into low- and high-revenue categories, and the difference in incumbent advertising behavior is examined. Fiona M. Scott Morton (2000) focuses on the determinants of generic entry in a dataset 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. Henry G. Grabowski and John M. 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 Catherine 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. One paper that has taken this approach is Vrinda Kadiyali's (1996) study of the market for film. It has been more common to provide indirect evidence of strategic investment by showing that investments affect future competition (which implies that investments will be strategic if firms are rational and aware of the effect on competition). One can think of Marvin B. Lieberman's (1987) discussion of the responses by incumbents in chemical industries to rivals' additions of capacity; Judith A. Chevalier's (1995a, 1995b) studies of the effect of capital restructuring on entry and exit and supermarket

[2] Some related papers, in addition to those discussed in the text, are Alison Masson and Robert L. Steiner (1985), Richard G. Frank and David S. Salkever (1992), Mats A. Bergman and Niklas Rudholm (2003), David Reiffen and Michael R. Ward (2005), and Tracy L. Regan (2008).

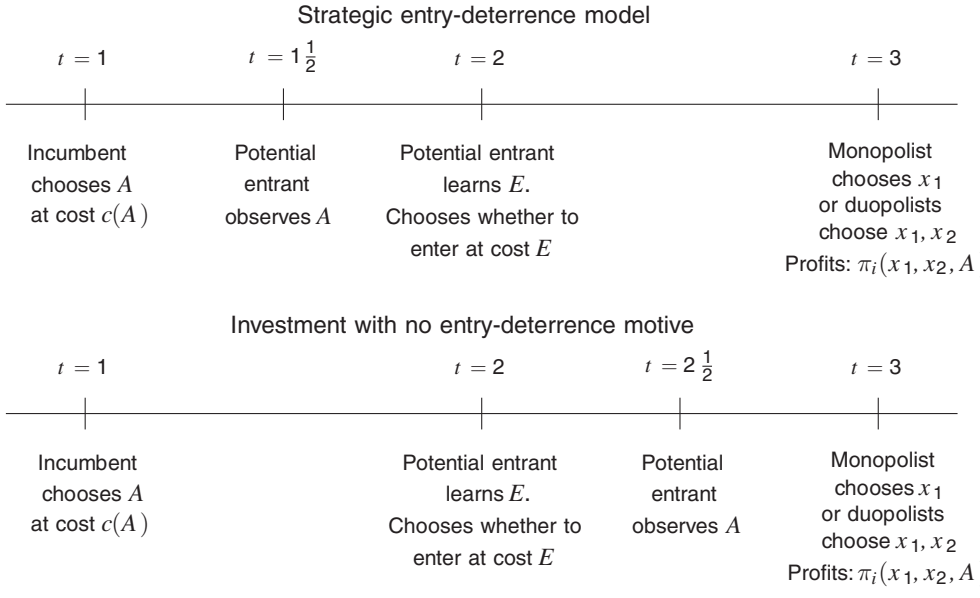

FIGURE 1. THE MODEL

pricing; and Scott Morton's (2000) discussion of the effects of advertising on entry as providing evidence of this sort. A third approach, taken by Austan Goolsbee and Chad Syverson (2008) 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.[3]

I. 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.

A. *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

[3] Other approaches have also been taken. Lieberman (1987) and Michael Conlin and Kadiyali (2006) examine whether excess capacity is systematically related to market structure in a cross section of markets. Robert Smiley (1988) reports evidence from surveys of firms. David J. Cooper, Susan Garvin, and John H. Kagel (1997) examines a limit-pricing model experimentally.

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

The one departure we have made from the way strategic entry-deterrence models are usually structured is that we assume that firm 2's entry cost E is stochastic with CDF $F(E)$ and so that firm 1 will not know for sure whether entry will occur when making its investment decision. We think this improves the theory and is necessary for empirical application.

B. *The Strategic Entry-Deterrence Incentive*

Let A_{ED}^* be the sequential equilibrium choice of A in this model. Investment is said to be "distorted" by the strategic entry-deterrence motive to the extent that A_{ED}^* differs from the investment level, A_{ND}^*, that would be chosen in the second game pictured in Figure 1.[4] In this game, firm 2 does not observe firm 1's investment level before the entry decision, so the investment cannot have a causal effect on entry. The nonstrategic benchmark A_{ND}^* can be thought of as reflecting what would happen if there was no entry-deterrence motive. It 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.

To understand the strategic entry-deterrence motive, it is useful to consider 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:

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

In the model with no 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. In equilibrium, firm 2 will assign probability one to firm 1 having chosen A_{ND}^*, so firm 1's expected profit is

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

[4] To avoid confusing people who know the literature, we should note that we have simplified the standard presentation to omit any mention of strategic entry accommodation. Our assumption that A is observed at $t = 2½$ in the model "with no strategic entry-deterrence motive" means that both of our models are models in which firms are assumed to recognize and react to incentives to accommodate entry (should it occur). Hence, any differences between A_{ED}^* and A_{ND}^* are entirely due to whether firms react to the incentive to deter entry which exists in the ED model but not in the ND model.

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

$$c'(A^*_{ND}) = F(\pi_2^{d*}(A^*_{ND}))\frac{\partial \pi_1^{d*}}{\partial A}(A^*_{ND}) + \big(1 - F(\pi_2^{d*}(A^*_{ND}))\big)\frac{\partial \pi_1^{m*}}{\partial A}(A^*_{ND})$$

$$c'(A^*_{ED}) = F(\pi_2^{d*}(A^*_{ED}))\frac{\partial \pi_1^{d*}}{\partial A}(A^*_{ED}) + \big(1 - F(\pi_2^{d*}(A^*_{ED}))\big)\frac{\partial \pi_1^{m*}}{\partial A}(A^*_{ED})$$

$$+ \big(\pi_1^{d*}(A^*_{ED}) - \pi_1^{m*}(A^*_{ED})\big)\frac{d\pi_2^{d*}}{dA}(A^*_{ED}) f\big(\pi_2^{d*}(A^*_{ED})\big).$$

The final term in the first-order condition for A^*_{ED} is the "strategic entry-deterrence" incentive. Because firm 1's profit is higher when in monopoly, it has an incentive to distort its investment to reduce firm 2's profit (which reduces the likelihood of entry).

Note that the strategic entry-deterrence incentive will often 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. In very small (or very large) markets, this likelihood will be small because the fixed entry costs will almost surely be much larger (smaller) than the duopoly profits. In intermediate-sized markets, there is a greater chance that the investment will have a pivotal effect on entry.

C. *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 (Targeted Advertising with Spillovers):

Consider a cross section of markets. Suppose that the ith market has a mass z_i of potential consumers, but that the markets are otherwise identical. Let A reflect the per-consumer 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 $z_i A$ on advertising in market i, a consumer of type θ receives utility $\theta\sqrt{2A} - p_1$ if he buys the (branded) good from firm 1 at price p_1, $(1/2)\,\theta\sqrt{2A} - p_2$ 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 $p_1 = (1/2)\sqrt{2A}$ and receives profit $(z/4)\sqrt{2A}$. The duopoly equilibrium prices are $p_1^* = (2/7)\sqrt{2A}$ and $p_2^* = (1/14)\sqrt{2A}$. Duopoly profits are $(8/49)z\sqrt{2A}$ and $(1/49)z\sqrt{2A}$.

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 $1/32$ at

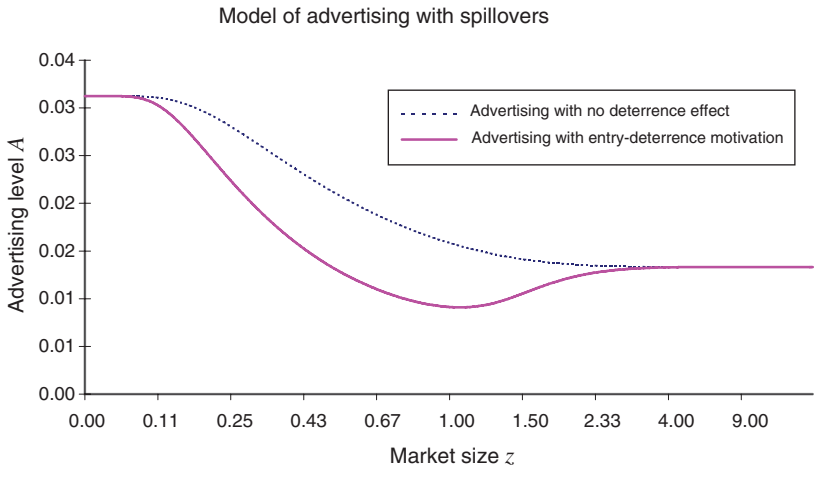

FIGURE 2. EQUILIBRIUM ADVERTISING LEVELS IN THE MODEL OF ADVERTISING WITH SPILLOVERS

Notes: The figure graphs the equilibrium advertising intensity in the model of Section IC, 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.

$z = 0$ to $32/2401$ in the limit as $z \to \infty$.[5] 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 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. The next section discusses the generality of this feature and the possibility of basing tests of strategic intent on it.

II. Monotonicity and Entry-Deterrence Motives

At a general level, our approach is to identify where two competing models make different predictions and then examine those predictions. 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 this model predicts that the investment-market size relationship will be monotone. Under those conditions, if the data are nonmonotone, then one can conclude that investments are influenced by the entry-deterrence motive (or that auxiliary assumptions of our propositions are violated).

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

A. *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., $\pi_2^{d*}(A,z)/\partial z > 0$.

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

DEFINITION 1: *The "direct effect" of z on A is* $F(\pi_2^*)\left(\partial^2\pi_1^{d*}/\partial z\partial A\right) + (1 - F(\pi_2^*)) \times \left(\partial^2\pi_1^{m*}/\partial z\partial A\right) - \left(\partial^2 c/\partial z\partial A\right)$.

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* $\left(\partial\pi_1^{d*}/\partial A\right) - \left(\partial\pi_1^{m*}/\partial A\right)$.

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 if firms are not influenced by entry-deterrence motives.[6]

PROPOSITION 1: *Let* $A_{ND}^*(z)$ *be the equilibrium investment level in the model of investment absent entry-deterrence motivations described above. Suppose* $(d\pi_2^{d*}/dz) > 0$.[7] *Then* $A_{ND}^*(z)$ *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.*[8]

[6] Recall that our model subsumes entry-accommodation motives.

[7] Note that this does involve an additional assumption. Earlier, we had assumed just that z was ordered so that $(\partial\pi_2^{d*}/\partial z) > 0$. Because $(d\pi_2^{d*}/dz) = (\partial\pi_2^{d*}/\partial z) + (\partial\pi_2^{d*}/\partial A)\,(dA/dz)$, 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 $(d\pi_2^{d*}/dz)$ before solving for dA/dz, it is easy to see that it suffices to 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.

[8] To 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

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 conditions 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 do not.

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 will not affect whether a relationship is monotonic, but other choices, like whether to use total or per capita advertising expenditure, can. Ideally, variables should be chosen so that the direct and competition effects will be of the same sign and the direct effects are not so strong as to make it implausible that strategic entry-deterrence motives will lead to nonmonotonicities.

Remark 3: The proposition says nothing about what happens if the direct and competition effects go in opposite directions. One can still look for evidence of strategic entry deterrence by looking for nonmonotonicities in such applications. One would, however, want to consider whether alternate explanations for the nonmonotonicity involving different nonstrategic effects dominating in different regions were 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.*

B. *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 more generally to build intuition for how Proposition 1 can be applied.

First, consider our previous example (Example 1) in which advertising increased consumer values for both the branded and generic good. In Section C, we showed graphically that advertising (measured as expenditure per potential consumer) was monotone decreasing in the market size. Why this occurred should now be easy to understand. The profit and cost functions satisfy the hypotheses of Corollary 1,

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.

so there is no direct effect. 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 preferences for the incumbent's product relative to the entrant's, and because the incumbent has a smaller market share in duopoly.)

We can get a simple example of a model for which nonstrategic investment may not be monotone by altering the advertising technology in Example 1. Example 1 assumed that advertising costs were proportional to the number of potential consumers. This might be a good model, for example, for an Internet advertising campaign targeted at consumers who searched for a disease-specific keyword. For some other forms of advertising, it would be more reasonable to assume that the cost of a campaign is independent of the number of potential consumers. For example, the cost of advertising a drug during the Super Bowl does not depend on the number of people who might benefit from the drug.

Example 2 (Broadcast Advertising with Spillovers):

Consider the model of Example 1 but alter the advertising technology so that the total cost of a campaign that increases each consumer's valuation by $\sqrt{2A}$ *is A regardless of the market size.*

In this model, the direct effect is positive: advertising is no more costly in larger markets and delivers greater benefits (holding market structure fixed). The competition effect remains negative. Hence, Proposition 1 does not apply. The fact that there are effects going in both directions does not imply that nonstrategic investment would be nonmonotone—it could be that the direct effect is so strong as to make investment increasing in z for all z . But we will get a nonmonotonic pattern for some entry cost distributions.

Another investment that we consider in our empirical section is product proliferation: firms may develop alternate "presentations" of a product, developing tablets with different strengths or oral liquid versions, for instance. Because pharmacists can only substitute a generic version of a drug if the presentation prescribed by the doctor is available in a generic, these investments may be particularly valuable when an entrant will only offer a strict subset of the incumbent's product line. The example below is motivated by the informal suggestion of Ellison et al. (1997) that a dual decision-maker process might be one way to account for patterns of cross-price elasticities between similar branded drugs, and between branded drugs and generic competitors.

Example 3 (Product Proliferation):

Let z_i be the mass of potential consumers for drug i. Suppose that if the incumbent invests A, then with probability $p(A)$ it invents an alternate presentation of the drug. Suppose that each potential consumer receives benefit v_s from the standard presentation. Suppose that the alternate presentation gives benefit $v_a > v_s$ to a fraction α of consumers and no benefit to other consumers. Suppose that consumer purchases are the result of a two-stage process. In the first stage, a doctor prescribes whichever presentation is better for each patient. In the second stage, consumers visit a

pharmacy, learn the price of the branded and generic version of their prescribed presentation (if a generic exists), and choose which firm to buy from, if any. Assume that a consumer's willingness to pay for a generic version of any presentation is a fraction θ of his or her willingness to pay for the branded version, with $\theta \sim \mathrm{U}[0,1]$. Assume that $\alpha v_a < v_s$ and that the entry cost distribution is such that the generic firm never produces the alternate presentation.

The direct effect is positive in this model: the benefits of the investment are proportional to z_i, whereas the costs are independent of z_i. The competition effect is also positive. A monopolist charges v_s and v_a for the two presentations, and therefore gets an incremental benefit of $z_i\alpha(v_a - v_s)$ from having the second presentation. If the generic firm enters and produces the standard product, it will set the price equal to $v_s/2$ and sell to half of the consumers. The incumbent's incremental benefit from having the second product is then $z_i\alpha(v_a - v_s/2)$.

Another model of advertising that would be natural for some applications (though probably not as important in our application) 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 a formal version:

Example 4 (Targeted Differentiating Advertising):
Suppose market i has a mass z_i of potential consumers with unit demands differentiated by a taste parameter θ which is uniformly distributed on $[-1,1]$. Suppose that after the incumbent in market i spends z_iA on advertising, a consumer of type θ receives utility $1 - p_1$ if he buys the good from the incumbent firm 1 at price p_1, $1 + \theta t(1 + \sqrt{2A}) - p_2$ if he buys from firm 2 at p_2, and zero if he makes no purchase.

With this specification duopoly profits are proportional to $1 + \sqrt{2A}$ and monopoly profits are independent of A. Hence, the competition effect is positive. Again, there is no direct effect. Hence, nonstrategic advertising levels will be monotone increasing. Here, the assumption that advertising is "targeted," i.e., that costs are proportional to the mass of potential consumers, is not critical. If advertising costs were instead independent of z_i , then there would simply be a positive direct effect that went along with the positive competition effect.

III. Econometric Issues

The theoretical results, above, suggest that one could provide evidence that investment levels reflect a strategic entry-deterrence motive, first by arguing that investment levels A should be monotone in a market size measure z absent entry-deterrence motives, and then by showing that they are actually nonmonotone. In this section, we discuss the econometric implementation of such a test.

A. *Standard Tests of Monotonicity*

Several econometric 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 Irène Gijbels et al. (2000); Subhashis Ghosal, Arusharka Sen, and Aad W. van der Vaart (2000); and Peter Hall and Nancy E. Heckman (2000). Hall and Heckman's (2000) 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 data points. 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

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

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.

B. *Our Implementation*

In this paper, we will test for monotonicity in two ways: one uses a slight modification of Hall and Heckman's (2000) test statistic; the other uses a new statistic we propose.

Our modification of Hall and Heckman's (2000) 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

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

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:

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

where W is a kernel weight matrix with elements w_{ij} reflecting differences in the z's, $\overline{W} = \left(W + W'\right)/2$, and $\hat{\sigma}^2 = \hat{\epsilon}'\hat{\epsilon}/N$. 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 this procedure is valid in our setting.[9] Therefore, we conduct simulations to help assess the validity of the procedure, and the power of the tests. We discuss these briefly in the Appendix.[10]

C. *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. 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 $A(r) \equiv E(A^*(z) \mid r)$. Whether looking for nonmonotonicity remains appropriate, depends on whether $A(r)$ inherits the monotonicity of $A^*(z)$ under the null. Standard results from incentive theory make it easy to give conditions under which this will hold. The conditional density $f(x \mid \theta)$ of a random variable is said to have the monotone likelihood ratio property (MLRP) in x if $f(x \mid \overline{\theta})/f(x \mid \underline{\theta})$ is monotone increasing in x whenever $\overline{\theta} > \underline{\theta}$. Under this assumption, we have

PROPOSITION 2: *Suppose $A^*(z)$ is monotone in z. Suppose the 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 (Paul R. 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 $A^*(z)$.

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 holds for

[9] The theoretical results of Hall and Heckman (2000) cannot be applied directly for a couple of 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.

[10] In unreported simulations, we also examined the power of the Ghosal, Sen, and van der Vaart (2000) and Gijbels et al. (2000) tests. We chose the two statistics we use because they were quicker to compute and/or more powerful in these simulations.

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.

D. *Endogeneity*

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

Endogeneity is a more serious concern. However, the fact that we are only interested in monotonicity (as opposed to obtaining consistent parameter estimates) will mitigate the concern in some circumstances. Formally, consider again a cross-section dataset containing investment levels A_i and a proxy r_i for the market size. Suppose that

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

$$r_i = r(z_i, A_i) + \eta_i,$$

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.

Proposition 3 has two primary limitations. One is that it only covers the case in which $A^*(z)$ is monotone increasing, not the case when $A^*(z)$ is monotone decreasing. The reader may be confused about how this can matter: what happens if one simply redefines the "investment" to be $-A$? The resolution to this is that "increasing" is meaningful because we are maintaining the assumption that the endogenous relationship is that r is increasing in A. An intuition for Proposition 3 is that when the two relationships go in the same direction, then A is increasing in r for two

reasons: A is larger because z is larger and $A^*(z)$ is monotone; and A is also larger in expectation because of the endogenous relationship in which larger values of A cause r to be larger.

A second limitation is that we have assumed that $\tilde{r}(z, \epsilon)$ satisfies two monotone likelihood ratio properties. If $r(z, A)$ is linear in its arguments, then one of these reduces to the assumption 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.

E. *IV Approaches*

In our application, we do not have compelling instruments. However, in other applications of our framework, it is possible that measurement error and endogeneity might addressed with an IV approach. The simplest example of such an approach would be to assume that the relationship between A and z had a known parametric form. For example, if one added the assumption that $A_i = \alpha_0 + \alpha_1 z_i + \alpha_2 z_i^2 + \ldots + \alpha_n z_i^n + \epsilon_i$ and had an instrument for the market size proxy, then one could estimate the parameters $\alpha_0, \ldots, \alpha_n$ via a simple IV regression (using polynomials of the instrument as instruments for the z^k). A polynomial is monotonic if and only if the coefficients belong to some set, and one could test the restriction that the coefficients belong to this set.

A natural idea for developing a nonparametric version of this approach would be to implement a nonparametric IV estimator as the first stage and then test whether the fitted function is monotonic. See Whitney K. Newey and James L. Powell (2003), Richard Blundell, Xiaohong Chen, and Dennis Kristensen (2007), and Chen and Demian Pouzo (2008), among others, for approaches to nonparametric IV estimation. Whether monotonicity tests can be validly applied to such estimates under some regularity conditions is beyond the scope of this paper.

IV. The Pharmaceutical Industry

In this section, we provide some background on the US 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.

A. *Industry Background*

Prior to 1984, all but the most popular drugs tended to retain their monopoly position in the US 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-1980s: the Waxman-Hatch Act of 1984 reduced

regulatory barriers to generic entry, and state laws mandating/allowing generic substitution by pharmacists boosted the market share of generic drugs.[11]

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 percent 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. For such drugs, generic entry is much less certain.

There are a number of investments that one could imagine pharmaceutical incumbents distorting in order to deter entry. The most obvious is advertising, which plays an important role in pharmaceutical markets—an oft-cited statistic by critics of the pharmaceutical industry is that more money is spent by the industry on marketing than on research and development. In the period we study, there were two main advertising channels: "detail" and "journal" advertising.[12] 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 is the placement of advertisements in medical journals and other publications read by doctors. Expenditures on detail advertising are typically much larger than expenditures on journal advertising.

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 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. When a drug is produced in many presentations, it would be more costly for an entrant to replicate the incumbent's full product line. A potential entrant can (and often does) choose to produce a strict subset of the set of presentations offered by the incumbent. This does, however, reduce subsequent profits. Rules on generic substitution vary from state to state, but they typically make it difficult for patients to substitute across presentations. For example, if a doctor has prescribed that a patient take one 100mg tablet per day, then the pharmacist may be prevented from dispensing 50mg tablets and instructing the consumer to take two tablets per day.

An additional instrument that firms might use to deter entry is pricing. Pharmaceutical manufacturers sell drugs to different consumers at different prices. Our data contain average wholesale prices for two classes of purchasers: hospitals and drugstores.[13]

B. *Data*

Our basic dataset includes 63 distinct chemical compounds, sold under 71 different brand names, that faced potential generic entry as the result of a patent or FDA

[11] See Grabowski and Vernon (1992).

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

[13] As discussed in Ellison and Christopher M. Snyder (2010), the former are typically substantially lower because of differences in the bargaining power of hospitals and drugstores.

Table 1—Variable Names

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

Note: The table describes the variables used in the analysis.

exclusivity expiration between 1986 and 1992. Details of the construction of our dataset appear in Appendix Section C.

The main variables used in the analysis are described in Tables 1 and 2. The first five variables in Table 2 are defined at the drug level. The mean of the *Entry*3*Yr* variable reflects that 37 of the 63 drugs experienced generic entry in the three-year window. The mean of the *Revenue*3 variable indicates that the average drug had annual revenues of $39.4 million.

*Detail*3 and *Journal*3 are average annual values of the advertising variables over the same three-year, pre-expiration period for which *Revenue*3 was computed. The values for the mean advertising ratios in Table 2 indicate that 1.4 percent of sales were spent on journal advertising and approximately 5 percent on detail advertising.[14] *PresHerf*3 is a Herfindahl-style measure of the degree to which revenues are concentrated in a small number of presentations in the three years prior to patent expiration.[15] 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.[16] The *Detail*3, *Journal*3, and *PresHerf*3 variables have 69 or 70 observations rather than 63 because we have defined them at the level of the brand name rather than at the level of the drug.[17]

[14] Our detailing data are in minutes. This calculation assumes a cost of $10 a minute.

[15] More precisely, it is the average over the three years of the annual concentration measure *PresHerf* described in Appendix Section C. For 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.

[16] Recall 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 *PresHerf*3 equal to one. The mean of *PresHerf*3 for the remaining drugs is still 0.46.

[17] Hence, 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.

TABLE 2—SUMMARY STATISTICS

Variable	Observations	Mean	SD
*Entry*3*Yr*	63	0.59	0.50
*Revenue*3	63	39,355	55,754
log(*Revenue*3)	63	9.40	2.00
HospFrac	63	0.21	0.30
Chronic	63	0.62	0.42
TherSubs	63	8.48	6.04
*Detail*3 / *Revenue*3	69	0.005	0.008
*Journal*3 / *Revenue*3	70	0.014	0.022
*PresHerf*3	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

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

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 1.9 percent and 1 percent above the rate of inflation.

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

C. *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), J. P. Bae (1997), 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 3 reports estimates from a probit regression of *Entry*3*Yr* on *Revenue*3 to verify that such a relationship exists in our data as well. The second column adds several other covariates to the regression. None except *Revenue*3 have a statistically significant effect on the likelihood of generic entry.[18]

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 large. To give some feel for the degree of heterogeneity in

[18] The 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 percent 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: Alan T. 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 (2010) 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.

Table 3—Entry versus Pre-Expiration Revenues

Dependent variable for probit is *Entry3Yr*		
Variable		
log(*Revenue*3)	0.70	0.76
	(0.17)	(0.20)
HospFrac		1.01
		(0.78)
Chronic		0.60
		(0.54)
log(*TherSubs*)		0.01
		(0.29)
Constant	−6.39	−7.60
	(1.62)	(1.94)
Observations	63	63
PseudoR^2	0.40	0.43

Notes: 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.

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.[19] 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.

V. Strategic Investment in Pharmaceuticals: Cross-Sectional Patterns in Incumbent Behavior

The three subsections of this section examine data on three potential strategic investments: detail advertising, journal advertising, and presentation proliferation. In each subsection, we first discuss what might be expected absent entry-deterrence motives, i.e., the likely "direct" and "competition" effects, and how behavior might be distorted to deter entry. We then present some descriptive regressions illustrating the cross-sectional patterns in the data, and comment on whether any nonmonotonicities are significant using the tests discussed in Section III.

[19] Note 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.

Table 4—Summary Statistics by Revenue Quintile

	Mean within revenue quintile				
Variable	Q 1	Q 2	Q 3	Q 4	Q 5
*Revenue*3	882	7,572	22,161	52,336	127,359
*Entry*3*Yr*	0.00	0.43	0.86	0.80	0.92
Observations	13	14	14	10	12

Note: 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.

A. *Detail Advertising*

Our Example 1 is intended to serve as a simple base model for the economics of detail advertising. It has two main assumptions. One is that advertising builds the market for a drug but provides spillover benefits to generic entrants. For two reasons, we believe such spillovers would be expected with pharmaceutical advertising. First, the FDA requires the content of pharmaceutical advertising to be based on the drug's therapeutic characteristics. Prescribing physicians will know that claims about therapeutic characteristics apply equally to generics (which are chemically identical). Second, even if the advertising bolstered physicians' opinions of the branded product relative to the generic, spillover benefits would accrue to generics at the dispensing stage where generic substitution is very common. The second main assumption of Example 1 is that advertising costs are proportional to the size of the patient population (more targeted than broadcast in nature). This assumption is most plausible for drugs prescribed by specialists: if condition A is twice as prevalent as condition B, then we would expect that there will be twice as many specialists treating condition A, and twice as many sales representative hours will be needed to visit the condition A doctors. The conclusion of Example 1 is that the advertising-to-market size ratio will be monotonically decreasing in the market size if firms are not influenced by strategic entry-deterrence motivations. Our cross-sectional analysis will examine advertising-to-sales ratios.[20] To the extent that sales is an appropriate proxy for market size, we would expect the advertising-to-sales ratio to be monotonically decreasing in the market size if advertising is not influenced by an entry-deterrence motive.[21]

Our graphical analysis of Example 1 illustrated the expected impact of an entry-deterrence motive. The form that "strategic entry deterrence" takes in this example is somewhat counterintuitive: the "strategic investment" is to reduce advertising. This reduction in advertising is a "playing dead" strategy, although it is not intended

[20] For 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.

[21] Note that the fact that the relationship should be monotone decreasing means that endogeneity is a potential concern. Note also that in addition to the concerns noted above, the use of sales instead of an ideal market size creates a division bias in the dependent variable. This would be another mechanism leading to a negative relationship between the left- and right-hand-side variables, but we have not identified any formal conditions under which the relationship would remain monotone decreasing.

TABLE 5—INCUMBENT BEHAVIOR VERSUS MARKET SIZE: LINEAR REGRESSIONS

Independent variable	Dependent variable		
	$\frac{Detail3}{Revenue3}$	$\frac{Journal3}{Revenue3}$	*PresHerf*3
log(*Revenue*3)	0.000	0.003	−0.069
	(0.001)	(0.002)	(0.016)
$(\log(Revenue3) - \overline{R})^2$	−0.0001	0.0000	−0.003
	(0.0002)	(0.0005)	(0.005)
Specialist	0.006	0.008	
	(0.009)	(0.026)	
Psych			−0.342
			(0.075)
Topical			−0.388
			(0.090)
Constant	−0.001	−0.018	1.290
	(0.008)	(0.023)	(0.168)
Observations	69	70	70
R^2	0.04	0.06	0.52

Notes: 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.

to fool generic competitors: reducing advertising simply reduces demand for the drug, which makes the market less attractive to entrants and thereby reduces the likelihood of generic entry. Recall that "intermediate-sized" means roughly where entry is uncertain, which in our case is around the second revenue quintile.

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

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

where $\overline{R}$ is the mean of log(*Revenue*3). 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 5. 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 6 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 *Detail*3/*Revenue*3 within each of the revenue quintiles. The most noteworthy value

Table 6—Incumbent Behavior versus Market Size: Quintile Means and Monotonicity Tests

	Variable mean for drugs in revenue quintile					Monotonicity test *p*-value	
Variable	Q 1	Q 2	Q 3	Q 4	Q 5	H-H test	E-E test
*Detail*3/*Revenue*3	0.0051	0.0013	0.0055	0.0084	0.0042	0.197	0.048
*Journal*3/*Revenue*3	0.011	0.005	0.011	0.024	0.018	0.080	0.227
*PresHerf*3	0.78	0.64	0.49	0.44	0.35	0.476	0.917

Notes: 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).

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. The means in the quintiles are, however, noisily estimated and sensitive to how the data is divided because there are a relatively small number of drugs that advertise heavily. To provide a formal test of whether the nonmonotonicity is significant without imposing arbitrary cutoffs, we carried out the tests described in Section B.[22] The sixth and seventh columns report estimated *p*-values for these two tests. The Ellison and Ellison (2000) test is statistically significant at the 5 percent level, but the Hall and Heckman (2000) test is not. We conclude that the level of noise and the size of the dataset make it hard for us to say whether apparent nonmonoticity in the table is significant.[23]

We should note that there are other factors that could be relevant to detail advertising that are not captured in Example 1, and these could provide alternate explanations for nonmonotonicy. For example, one could argue that the appropriate model is more like that of Example 2 and the detail-to-sales ratio declines from small- to intermediate-sized markets because of the competition effect (the benefits are smaller when the incumbent will only capture a share of the gains), but then increases in larger markets because of a positive direct effect, e.g., fixed costs are such that only very popular drugs can be advertised to nonspecialist doctors. Indeed, we think a nice feature of our framework is that it makes it easy to think about whether other factors would change the qualitative conclusions. For example, one could argue that detail advertising may also serve to differentiate branded products from generics along the lines of Example 4. Our understanding is that this effect is not very important in pharmaceuticals, but it is an effect that goes in the opposite direction, and one could think about whether it might be a plausible alternate explanation if advertising-to-sales ratios were increasing in some region.

[22] In these and all other monotonicity tests, we used a Probit transformation of the log(*Revenue*3) variable as the z variable (using the coefficients in the first column of Table 3). This makes the range of z a large subset of [0,1] and seemed desirable because the kernel regression approximations to the function are derived from pooling observations with similar probabilities of generic entry. We set the smoothing parameter for the Ellison and Ellison (2000) style test to be equal to 0.2 and used a parameter for the Hall and Heckman (2000) test corresponding to subregressions of 15 observations.

[23] Note that we made no assumptions about the direction of entry-deterring activity in the design of our test. We could have assumed, based on our reasoning above, that the direction of entry deterrence would be to decrease advertising, and designed a test to look specifically for deviations in that direction. Such a test would be expected to have higher power.

B. *Journal Advertising*

Our second potential tool for strategic entry deterrence is journal advertising. We think of journal advertising as a less natural setting in which to use monotonicity tests as a test of strategic intent—the market may be better described by Example 2 than Example 1. The competition effect should be negative as before: doctors will realize that information in medical journal advertisements applies also to generics, and there will be substitution at the pharmacy level. But Proposition 1 won't apply, because the direct effect should be positive: the cost per potential patient should be decreasing in the size of the pool of potential patients.[24] The strategic distortion would be the same as for detail advertising: firms should reduce advertising in intermediate-sized markets to reduce the attractiveness of the market to potential entrants.

The second column of Table 5 reports estimates from a regression of *Journal*3/*Revenue*3 on $\log(Revenue3)$, $(\log(Revenue3) - \overline{R})^2$ 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 6 indicate that journal advertising is also lowest in the second quintile. The results are similar to what we found for detail advertising, but with the role of the two tests reversed: the Hall and Heckman (2000) test rejects monotonicity at the 8 percent level while the Ellison and Ellison (2000) test does not.

We would summarize these results by saying that they suggest that there may again be some nonmonoticity in the relationship between journal advertising and market size, although the evidence here is even weaker. The form of the nonmonotonicity is consistent with what one would expect from an entry-deterrence model. For the journal advertising application, however, the basic economics of the problem had suggested to us from the start that a nonstrategic story could also be given: it could be that the competition effect dominates in small markets whereas economies of scale drive the cross-sectional pattern in large markets.

C. *Presentation Proliferation*

The final incumbent action we examine here is presentation proliferation. Example 3 reflects two considerations that we suspect are important. First, we think there would be a strong positive direct effect: firms will invest more in developing new presentations when the market is larger because the costs of developing are mostly fixed, whereas the benefits scale with the size of the market. Second, a factor that may make the competition effect positive is that the incremental benefits of a new presentation are largest when it gives the incumbent a monopoly niche that it otherwise would not have had. Hence, our presumption is that investment in presentation proliferation would likely be monotone increasing in market size absent entry-deterrence motives. Our dependent variable, *PresHerf*3, is a Herfindahl-style

[24] The situation would be more like Example 1 if ads were placed primarily in specialty journals that only reached doctors who specialize in treating some condition.

measure which is smaller when the number of presentations is larger. Hence, we would expect that it would be monotone decreasing in market size.[25]

Again, although not necessary for our test, it is useful to note that 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 action 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) and would be reflected in a reduction in our Herfindahl-style measure, *PresHerf*3, in intermediate-sized markets.[26] An aspect of this application that makes it less than ideal is that the direct effect may be very strong. Strategic investment will only lead to nonmonotonicity if the strategic incentive is strong enough to outweigh the direct effect.

The third column of Table 5 reports coefficient estimates from a regression of *PresHerf*3 on log(*Revenue*3), $(\log(\textit{Revenue}3) - \overline{R})^2$, and two control variables. There is clearly a strong relationship between presentation proliferation and revenues: the coefficient on log(*Revenue*3) is negative and highly significant. The quadratic term yields no evidence of nonmontonicity. The control variables, *Psych* and *Topical*, are also highly significant.[27]

Table 6 similarly gives no indication of a nonmonotone pattern. The quintile means are monotonically decreasing. And the two monotonicity tests indicate that there is nothing in the data to suggest a nonmonotonic pattern: the *p*-values are 0.48 and 0.92.

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 appears to be a case where our approach will have limited power.

VI. Strategic Investment in Pharmaceuticals: A Differences-in-Differences Approach

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

One can think of the analysis in the previous section 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

[25] Note that the fact that we expect *PresHerf*3 to be decreasing means that endogeneity is not a concern. What matters is the revenue-enhancing investment which we expect here to be increasing.

[26] An effect that could go in the other direction is that presentation proliferation can make the market more attractive to entrants if it expands the market and the cost of copying the added presentations is not large.

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

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 its slope may overwhelm the nonomonotonicity 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 it is not steeply sloped, 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 time t' when 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 for two reasons: the potentially steeply sloped nonstrategic term has been eliminated, and the error variance may be smaller. (If the nonstrategic term really completely drops out as above, then the function is identically zero absent an entry-deterrence motive and the framework can be applied regardless of whether nonstrategic investment is monotone.)

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.[28]

In the subsections that follow, we examine whether changes in each of the potential "strategic investments" are nonmontone in market size. We drop all

[28] In the case of the advertising variables, which are available to us at monthly frequency, we do this by comparing the 12 months prior to patent expiration with the preceding 24 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.

Table 7—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
*Detail*3	0.50	0.11	0.33	0.38	0.38	0.824	0.221
	(4)	(9)	(12)	(13)	(13)		
*Journal*3	0.00	0.43	0.00	0.14	0.23	0.079	0.066
	(2)	(7)	(12)	(14)	(13)		
PresHerf	0.33	0.42	0.38	0.50	0.62	0.082	0.087
	(6)	(12)	(13)	(14)	(13)		
DPrice	0.70	0.58	0.75	0.54	0.92	0.356	0.200
	(10)	(12)	(12)	(13)	(13)		
HPrice	0.50	0.50	0.54	0.77	0.73	0.564	0.678
	(8)	(12)	(13)	(13)	(11)		

Notes: 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.

observations for there is zero advertising in both years from the change-in-advertising analyses, and drugs that are only ever offered in a single presentation from the change-in-presentation-proliferation analyses. Our primary dependent variable is an indicator variable for the behavior having increased or decreased. This empirical strategy makes the comparisons as simple as possible and eliminates problems due to outliers and heteroskedasticity.

A. *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 entrants. 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 7 reports the fraction of drugs in each revenue quintile that increased their detail/sales ratio in the year prior to patent expiration (relative to the detail/sales ratio for the two prior years). 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 36 months prior to patent expiration. Note that a pattern roughly consistent with the entry-deterrence model is visible: almost all of the drugs in the second quintile are decreasing detailing as patent expiration approaches; and the fraction of drugs that are decreasing detail advertising is higher here than in any other quintile. The use of a binary dependent variable makes this analysis less sensitive to outliers than our cross-section analysis, but the number of observations is now even smaller, and in particular there are only four drugs in the first quintile that do any detail advertising, which would make it hard for any pattern to be highly significant. The nonparametric tests each indicate that the pattern here is not statistically significant.

We conclude that when we examine detail advertising changes prior to patent expiration instead of levels, we cannot provide significant evidence of strategic entry deterrence. The form of the pattern is what one would expect under a strategic

entry-deterrence theory, but we have too few observations of real changes to say that anything is significant.

B. *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 7 indicates that reducing journal advertising is most common among firms in the third (and first) revenue-quintile. The nonparametric tests are consistent in that they both regard the nonmonotonicity as significant at the 6 percent to 8 percent level. But given that the form of the nonmonotonicity is not what would be expected given the theory, we would not regard this finding as evidence of strategic entry deterrence.

On the whole, journal advertising turns out to be a less interesting application. The fact that the direct and competition effects go in opposite directions means that nonstrategic explanations for nonmonotonicities will be plausible. Our cross-section analysis had provided some weak evidence that journal advertising is lower in intermediate-sized markets, as one would expect in an entry-deterrence model. Looking at changes in journal advertising as patent expiration approaches does not provide any additional support.

C. *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. We noted earlier that the cross-sectional relationship between market size and presentation proliferation is dominated by a strong trend, which is consistent with what one would expect given that the costs of developing presentations are largely fixed costs. Accordingly, one might expect our differences-in-differences approach to have more incremental benefit here.

The quintile means, reported in the third main row of Table 7, suggest the possibility of nonmonotonicity. The two nonparametric tests both indicate that the departure from monotonicity is significant at the 8 percent to 9 percent level. However, the magnitudes of the nonmonotonicities are not large, and the pattern does not match well with what one would expect from a strategic entry-deterrence model.

We conclude that looking at how firms change their presentation proliferation in the year prior to patent expiration yields some weak evidence of nonmonotonic behavior. But the magnitudes of the nonmonotonicities are not large and the pattern is such that it cannot be regarded as compelling evidence of firms' being influenced by a strategic entry-deterrence motive.

D. *Pricing*

Pricing is another case in which looking at changes in behavior could have a large 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.[29] Looking at whether firms are increasing or decreasing prices as patent expiration approaches, in contrast, is both simple and sensible.

The challenge of constructing a theory of prices as an entry-deterring investment is the lack of a physical link between the past and future prices: setting a low price today has no effect on entry if entrants expect the incumbent to jump to the static duopoly price as soon as entry occurs. 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.[30]

Variants of some of these stories would be plausible for the pharmaceutical application. One possibility is that current and future prices may be linked for political economy reasons: pharmaceutical firms come under substantial scrutiny when they raise prices and hence setting a low price today will make it more costly for a firm to set a higher price if a generic subsequently enters. Our discussions with industry sources suggest that the textbook signaling model of entry deterrence is not highly plausible: generic firms are well-informed about both prices and revenues (and indeed have all the same data we have). However, one could easily adapt a signaling model to make it more plausible. Generic firms will not be well informed about price elasticities, so 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 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.[31]

The fourth row of Table 7 reports the frequency with which firms in each quintile are raising prices charged to drugstores in the year before patent expiration.[32] 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.

[29] What 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.

[30] Among the early papers in this literature are Milgrom and John Roberts (1982), Fudenberg and Tirole (1983, 1986), Joseph E. Harrington, Jr. (1986), and Klemperer (1987).

[31] Incumbents 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 Zvi Griliches and Iain 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.

[32] To 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.

VII. 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 nonmonotonic patterns that are consistent with what one would expect to see if incumbents' actions were influenced by a desire to deter generic entry. In our cross-sectional analysis, we noted that detail and journal advertising were relatively low in intermediate-sized markets, which is what one would expect if firms strategically let a marginal market decay to make it less attractive to generic entrants. Our analysis of changes in detail advertising as patent expiration approached found a similar pattern.

We would be hesitant, however, to describe these results as indicative of entry-deterring behavior without providing a number of caveats. First and foremost, the nonmonotonicities are only marginally significant when one assesses them using nonparametric tests. Second, a variety of alternative nonstrategic explanations could be given—indeed we would argue that a benefit of our framework is that it makes it easy to think through various alternatives. Third, we have found just a few significant nonmonotonicities relative to the number of tests we have carried out on several potential strategic investments. Accordingly we would be interested to see further analyses of richer datasets that might help assess whether the limited number of significant results is because there is not much strategic behavior, or because our nonparametric tests will have limited power to detect moderate effects given the limited sample size and the demanding benchmark of significance in a fully nonparametric test.

From a practical perspective, it should be pointed out that the possibly entry-deterring 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, if incumbents have indeed figured out the potential gains from altering future competitive conditions on these unimportant drugs, then one would assume that firms are also strategically sophisticated with regard to their more important products. However, our approach cannot provide any direct evidence on whether blockbuster patent holders are behaving strategically.

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 three 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.[33]

[33] See Leemore S. Dafny (2005) for an interesting application to hospital markets.

We hope also that our paper may 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.

Appendix

A. *Proof of Proposition 1*

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

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

Differentiating with respect to z gives

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

where we have written $d\pi_2^{d*}/dz$ 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 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 $d\pi_2^{d*}/dz > 0$, 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.

TABLE A1—SIMULATION STUDY OF SIZE AND POWER OF MONOTONICITY TESTS

Data generating process	Rejections with 5 percent 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

Note: The table presents rejection rates for the Hall and Heckman (2000) and Ellison and Ellison (2000) style monotonicity tests when applied to five data-generating processes with 5 percent critical values obtained from a bootstrap procedure.

B. *Monte Carlo Study of Monotonicity Tests*

Table A1 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 500 simulated datasets, ran our bootstrap procedure to find estimated 5 percent critical values, and then compared the value of the test statistic on the simulated dataset to the estimated critical value.[34] The simulated datasets had 100 data points with uniform draws on $[0,1]$. The first column reports rejection rates (based on 5 percent critical values) for the Hall and Heckman (2000) style test with smoothing parameter $m = 15$. The second column reports rejection rates for the Ellison and Ellison (2000) 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 $\Pr\{A_i = 1 \,|\, z_i\} = 0.25 + 0.5z_i$. The rejection rates are usually around 5 percent 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 $\Pr\{A_i = 1 \,|\, z_i\} = 0.25 + 2z_i(1 - z_i)$. The rejection rates in these simulations range from 26.8 percent to 47.2 percent. 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.

[34] We used 1,000 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.

C. *Data Description*

Our basic dataset includes 63 distinct chemical compounds that faced potential generic entry as the result of a patent or FDA exclusivity expiration between 1986 and 1992.[35, 36]

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: *Revenue*3 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 36 months prior to patent expiration, the month of patent expiration, and the 11 subsequent months.

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.[37] *PresHerf* will be large in markets where a small number of

[35] Our dataset 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.

[36] These 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 task 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*.

[37] Defining 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

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.[38]

We obtained information on drug characteristics and whether generic entry did in fact occur from several other sources.[39] The primary variable we will use to study entry, *Entry*3*Yr*, 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.[40] *Chronic* is set to zero for drugs which treat an acute condition, and to one for drugs which treat a chronic condition.[41] *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*.

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.

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

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

[40] Caves, Whinston, and Hurwitz (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.

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

D. *List of Drugs Used in the Study*

Generic name	Brand names	Expiration	*Revenue*3	*Entry*3*Yr*
Albuterol	Proventil Ventolin	1989	172,952	1
Amiodarone Hydrochloride	Cordarone	1990	11,283	0
Amoxapine	Asendin	1989	18,306	1
Atenolol	Tenormin Tenoretic	1991	301,311	1
Auranofin	Ridaura	1992	9,766	0
Baclofen	Lioresal	1986	12,033	1
Betamethasone	Celestone	1986	8,226	0
Bretylium 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 Mycelex Lotrimin	1989	55,283	0
Colestipol Hydrochloride	Colestid	1989	7,089	0
Cromolyn Sodium	Nasalcrom Intal	1989	49,640	1
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 Sinequan	1986	55,059	1
Dronabinol	Marinol	1990	1,454	0
Enflurane	Ethrane	1987	19,337	1
Fenoprofen calcium	Nalfon	1988	49,538	1
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 Sanorex	1990	2,001	0
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

(*Continued*)

Generic name	Brand names	Expiration	*Revenue*3	*Entry*3*Yr*
Stanozolol	Winstrol	1989	541	0
Sulfasalazine	Azulfidine	1988	9,181	1
Sulindac	Clinoril	1990	164,545	1
Timolol Maleate	Blocadren Timoptic	1989	114,148	1
Tolmetin sodium	Tolectin	1990	48,654	1
Tretinoin	Retin A	1990	61,167	0
Trilostane	Modrastane	1989	24	0
Verapamil Hydrochloride	Isoptin Calan	1986	56,494	1

REFERENCES

Bae, J. P. 1997. "Drug Patent Expirations and the Speed of Generic Entry." *Health Services Research*, 32(1): 87–101.

Bergman, Mats A., and Niklas Rudholm. 2003. "The Relative Importance of Actual and Potential Competition: Empirical Evidence from the Pharmaceuticals Market." *Journal of Industrial Economics*, 51(4): 455–67.

Blundell, Richard, Xiaohong Chen, and Dennis Kristensen. 2007. "Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves." *Econometrica*, 75(6): 1613–69.

Bulow, Jeremy I., John D. Geanakoplos, and Paul D. Klemperer. 1985. "Multimarket Oligopoly: Strategic Substitutes and Complements." *Journal of Political Economy*, 93(3): 488–511.

Caves, Richard E., Michael D. Whinston, and Mark A. Hurwitz. 1991. "Patent Expiration, Entry, and Competition in the U.S. Pharmaceutical Industry." *Brookings Papers on Economic Activity: Microeconomics*, 1991: 1–66.

Chen, Xiaohong, and Demian Pouzo. 2008. "Estimation of Nonparametric Conditional Moment Models with Possibly Nonsmooth Moments." Yale University Cowles Foundation Discussion Paper 1650.

Chevalier, Judith A. 1995a. "Capital Structure and Product-Market Competition: Empirical Evidence from the Supermarket Industry." *American Economic Review*, 85(3): 415–35.

Chevalier, Judith A. 1995b. "Do LBO Supermarkets Charge More? An Empirical Analysis of the Effects of LBOs on Supermarket Pricing." *Journal of Finance*, 50(4): 1095–1112.

Conlin, Michael, and Vrinda Kadiyali. 2006. "Entry-Deterring Capacity in the Texas Lodging Industry." *Journal of Economics and Management Strategy*, 15(1): 167–85.

Cooper, David J., Susan Garvin, and John H. Kagel. 1997. "Signalling and Adaptive Learning in an Entry Limit Pricing Game." *RAND Journal of Economics*, 28(4): 662–83.

Dafny, Leemore S. 2005. "Games Hospitals Play: Entry Deterrence in Hospital Procedure Markets." *Journal of Economics and Management Strategy*, 14(3): 513–42.

Ellison, Glenn, and Sara Fisher Ellison. 2000. "A Simple Framework for Nonparametric Specification Testing." *Journal of Econometrics*, 96(1): 1–23.

Ellison, Sara Fisher, Iain Cockburn, Zvi Griliches, and Jerry Hausman. 1997. "Characteristics of Demand for Pharmaceutical Products: An Examination of Four Cephalosporins." *RAND Journal of Economics*, 28(3): 426–46.

Ellison, Sara Fisher, and Christopher M. Snyder. 2010. "Countervailing Power in Wholesale Pharmaceuticals." *Journal of Industrial Economics*, 58(1): 32–53.

Ellison, Sara Fisher, and Catherine Wolfram. 2006. "Coordinating on Lower Prices: Pharmaceutical Pricing under Political Pressure." *RAND Journal of Economics*, 37(2): 324–40.

Frank, Richard G., and David S. Salkever. 1992. "Pricing, Patent Loss and the Market for Pharmaceuticals." *Southern Economic Journal*, 59(2): 165–79.

Fudenberg, Drew, and Jean Tirole. 1983. "Learning-by-Doing and Market Performance." *Bell Journal of Economics*, 14(2): 522–30.

Fudenberg, Drew, and Jean Tirole. 1984. "The Fat-Cat Effect, the Puppy-Dog Ploy, and the Lean and Hungry Look." *American Economic Review*, 74(2): 361–66.

Fudenberg, Drew, and Jean Tirole. 1986. "A "Signal-Jamming" Theory of Predation." *RAND Journal of Economics*, 17(3): 366–76.

Ghosal, Subhashis, Arusharka Sen, and Aad W. van der Vaart. 2000. "Testing Monotonicity of Regression." *Annals of Statistics*, 28(4): 1054–82.

Gijbels, Irène, Peter Hall, M. C. Jones, and Inge Koch. 2000. "Tests for Monotonicity of Regression Mean with Guaranteed Level." *Biometrika*, 87(3): 663–73.

Goolsbee, Austan, and Chad Syverson. 2008. "How Do Incumbents Respond to the Threat of Entry? Evidence from the Major Airlines." *Quarterly Journal of Economics*, 123(4): 1611–33.

Grabowski, Henry G., and John M. Vernon. 1992. "Brand Loyalty, Entry, and Price Competition in Pharmaceuticals after the 1984 Drug Act." *Journal of Law and Economics*, 35(2): 331–50.

Grabowski, Henry, and John Vernon. 1996. "Longer Patents for Increased Generic Competition in the US: The Waxman-Hatch Act after One Decade." *PharmacoEconomics*, 10: 110–23.

Griliches, Zvi, and Iain Cockburn. 1994. "Generics and New Goods in Pharmaceutical Price Indexes." *American Economic Review*, 84(5): 1213–32.

Hall, Peter, and Nancy E. Heckman. 2000. "Testing for Monotonicity of a Regression Mean by Calibrating for Linear Functions." *Annals of Statistics*, 28(1): 20–39.

Harrington, Joseph E., Jr. 1986. "Limit Pricing When the Potential Entrant Is Uncertain of Its Cost Function [Limit Pricing and Entry under Incomplete Information: An Equilibrium Analysis]." *Econometrica*, 54(2): 429–37.

Hurwitz, Mark A., and Richard E. Caves. 1988. "Persuasion or Information? Promotion and the Shares of Brand Name and Generic Pharmaceuticals." *Journal of Law and Economics*, 31(2): 299–320.

Kadiyali, Vrinda. 1996. "Entry, Its Deterrence, and Its Accommodation: A Study of the U.S. Photographic Film Industry." *RAND Journal of Economics*, 27(3): 452–78.

Klemperer, Paul. 1987. "The Competitiveness of Markets with Switching Costs." *RAND Journal of Economics*, 18(1): 137–50.

Lieberman, Marvin B. 1987. "Postentry Investment and Market Structure in the Chemical Processing Industries." *RAND Journal of Economics*, 18(4): 533–49.

Masson, Alison, and Robert L. Steiner. 1985. "Generic Substitution and Prescription Drug Prices: Economic Effects of State Drug Product Selection Laws." Federal Trade Commission Staff Report.

Milgrom, Paul R. 1981. "Good News and Bad News: Representation Theorems and Applications." *Bell Journal of Economics*, 12(2): 380–91.

Milgrom, Paul, and John Roberts. 1982. "Limit Pricing and Entry under Incomplete Information: An Equilibrium Analysis." *Econometrica*, 50(2): 443–59.

Newey, Whitney K., and James L. Powell. 2003. "Instrumental Variable Estimation of Nonparametric Models." *Econometrica*, 71(5): 1565–78.

Regan, Tracy L. 2008. "Generic Entry, Price Competition, and Market Segmentation in the Prescription Drug Market." *International Journal of Industrial Organization*, 26(4): 930–48.

Reiffen, David, and Michael R. Ward. 2005. "Generic Drug Industry Dynamics." *Review of Economics and Statistics*, 87(1): 37–49.

Rosenthal, Meredith B., Ernst R. Berndt, Julie M. Donohue, Richard G. Frank, and Arnold M. Epstein. 2002. "Promotion of Prescription Drugs to Consumers." *New England Journal of Medicine*, 346(7): 498–505.

Scott Morton, Fiona M. 1999. "Entry Decisions in the Generic Pharmaceutical Industry." *RAND Journal of Economics*, 30(3): 421–40.

Scott Morton, Fiona M. 2000. "Barriers to Entry, Brand Advertising, and Generic Entry in the US Pharmaceutical Industry." *International Journal of Industrial Organization*, 18(7): 1085–1104.

Smiley, Robert. 1988. "Empirical Evidence on Strategic Entry Deterrence." *International Journal of Industrial Organization*, 6(2): 167–80.

Sorensen, Alan T. 2000. "Equilibrium Price Dispersion in Retail Markets for Prescription Drugs." *Journal of Political Economy*, 108(4): 833–50.

Tirole, Jean. 1988. *The Theory of Industrial Organization*. Cambridge, MA: MIT Press.

This article has been cited by:

1. Gautier Duflos, Frank R. Lichtenberg. 2011. Does competition stimulate drug utilization? The impact of changes in market structure on US drug prices, marketing and utilization. *International Review of Law and Economics* . [CrossRef]