Positive Spillovers and Free Riding in Advertising of Prescription Pharmaceuticals: The Case of Antidepressants

Bradley T. Shapiro

University of Chicago

Exploiting the discontinuity in advertising along the borders of television markets, I estimate that television advertising of prescription antidepressants exhibits significant positive spillovers on rivals' demand. I apply this identification in a demand model, where estimated parameters indicate significant and persistent spillovers driven by market expansion. Using the demand estimates to calibrate a stylized supply model, I explore the consequences of the positive spillovers on firm advertising choice. Compared with a competitive benchmark in which firms optimally free ride, simulations suggest that a category-wide advertising cooperative would produce a significant increase in total advertising.

I. Introduction

How does television advertising affect the consumer choice problem? After a consumer watches a commercial, internalizes its message, and decides a product is desirable, she must take further action to obtain the product.

I would like to extend my gratitude to Nancy Rose, Ernst Berndt, Stephen Ryan, Andrew Ching, Sara Fisher Ellison, Glenn Ellison, Günter Hitsch, Greg Leiserson, Sarah Moshary, Jesse Shapiro, Xiao Yu May Wang, editor Ali Hortaçsu, and two anonymous referees for their helpful suggestions. This paper also benefited from conversations with Paul Snyderman and Len Tacconi. I thank Cindy Halas at IMS Health, Jackie Allen and Leslie Walker at Kantar Media Intelligence, and LuAnn Patrick and Patrick Angelastro at ImpactRx for their help with data resources. This research was supported by the National Institute on Aging, grant T32-AG000186, the National Science Foundation Graduate Research Fellowship under grant 1122374, and the Beatrice Foods Co. Faculty Research Fund at the University of Chicago Booth School of Business. All mistakes are my own. Data are provided as supplementary material online.

Electronically published January 11, 2018
[*Journal of Political Economy*, 2018, vol. 126, no. 1]

With groceries, she must go to the supermarket. With many consumer products, a computer with internet will allow the consumer to make the purchase. With prescription drugs, the consumer must go to the physician to obtain a prescription and then to the pharmacy to purchase the drug. With many steps between the advertising incidence and purchase, at some stage of the process, the consumer might well choose a product different from the one advertised. This decision may be due to difficulty in remembering advertisements, agency problems in obtaining products, or simply that advertising convinces a consumer to go to a retailer, computer, or physician. In short, an advertisement could affect the choice process without leading the consumer to buy the advertised product.

In this paper, I identify the existence of positive spillovers of television advertising in the market for antidepressants. Given these spillovers, I construct and estimate a demand model that allows such spillovers. To quantify the potential size of the incentive effects of spillovers on firm behavior, I conduct a supply-side analysis supposing that the firms are able to decide advertising jointly, and I compare this outcome to a benchmark competitive outcome in which firms optimally free ride in the antidepressant market.

Branded television advertising of prescription drugs is contentious, and many have condemned it as inefficiently distorting prescriptions toward the advertised products. In fact, such advertising is legal in only two countries: New Zealand and the United States. In light of the controversy, understanding the impact of these advertisements is important. In particular, understanding spillovers is crucial to regulators, firms, and econometricians. From a regulatory perspective, the Food and Drug Administration (FDA) regulates the content of advertisements. To the extent that advertising content is made more informative and less brand-specific, content regulation could exacerbate spillovers. Firms may lose individual incentives to advertise as spillovers intensify. Diminished advertising incentives could be either good or bad for social welfare depending on whether category expansion is a public good or a public bad. However, it is an important consideration for the regulator in either case. From a firm strategy perspective, understanding possible channels for revenue improvement is vital. Although cooperation is often difficult to enforce and noncontractible because of antitrust laws, advertising cooperatives are precedented in other industries, such as for orange juice, milk, and beef. Finally, from a technical perspective, failure to model spillovers in advertising can distort estimated parameters, leading to incorrect inferences about supply and demand.

Previous research incorporating advertising into demand analysis has frequently treated advertising of a product as affecting its probability of being in the choice set (Goeree 2008) or has incorporated advertising into a production of goodwill that enters directly into the utility function (Dubé, Hitsch, and Manchanda 2005). However, such specifications also typically exclude the possibility of positive spillovers of advertising onto rivals. While this exclusion eliminates the complexity of modeling behavior in the presence of possible free riding, it may lead the researcher to miss important strategic considerations. When deciding how much to advertise, firms do not internalize the benefit they provide to other firms and have an incentive to free ride on their rivals' advertising efforts. Understanding these considerations is important for marketing decision makers as well as policy makers potentially seeking to regulate advertising.

Prescription drugs in general, and antidepressants in particular, have many characteristics that facilitate positive spillovers in television advertising. First, the FDA regulates what firms can and cannot say in advertisements. Although the name of the product is typically prominently displayed throughout the commercial, most of the time in each commercial is spent explaining the ailment, the mechanism of action of the drug, and its side effects. When several therapeutic products are available, those treating the same ailments tend to share common characteristics. A consumer might remember all the things being said but forget the name of the product. Agency problems further disrupt this link. A consumer must see a doctor to get a prescription. A physician might have different preferences or opinions about which drugs, if any, work best for a given condition or patient. The advertisement may lead a patient to the physician, but the physician remains the ultimate arbiter of whether and what to prescribe.

My strategy for evaluating the extent of positive spillovers in advertising for antidepressants proceeds in three steps. First, I use discrete television market borders to determine the extent to which advertising does affect rival demand, positively or negatively. Next, I construct and estimate a model of the antidepressant market, allowing advertising to have positive spillovers on demand of horizontally differentiated products, a feature that typical discrete-choice specifications exclude. Positive spillovers are allowed but not imposed by the model. Further, I find that not using the border discontinuity and assuming that advertising choices are exogenous leads the researcher to overstate the long-run effectiveness of advertising as well as understate the extent of the positive spillover. Next, I test whether firm behavior is consistent with free riding off of rival advertising efforts. I find that firms advertise less and less often when positive shocks to rival advertising occur in a given market. Finally, given estimates of the demand effects and an assumed marginal cost of advertising, I quantify the importance of free riding by simulating a stylized supply model and compare a benchmark competitive outcome whereby firms optimally free ride with a scenario with a cooperative that sets advertising for the entire industry.

Whereas most models incorporating advertising into demand have not allowed for positive spillovers, studies of direct-to-consumer (DTC) advertising in pharmaceuticals with varying credibility of identification strategies have shown some evidence that cross-advertising elasticities could be positive, but results have been mixed. In contrast to the demand analyses mentioned above that typically do not allow for cross-advertising elasticities to be positive, these studies tend to find patterns that are consistent with spillovers rather than allowing for spillovers within a demand model. In particular, Wosinska (2002), Donohue and Berndt (2004), and Iizuka and Jin (2005, 2007) find very small estimates of advertising effects on market shares, conditional on being in the market, and conclude that advertising might be exhibiting positive spillovers, though spillovers are neither directly modeled nor tested. Donohue and Berndt (2004) and Wosinska (2005) find that advertising has positive spillover effects onto drug compliance and duration of treatment. Sinkinson and Starc (forthcoming) find some evidence of brand advertising spillovers onto generics in the statin category. Other studies find that advertising drives consumers to the doctor (Iizuka and Jin 2005) or has class-level effects (Rosenthal et al. 2003; Avery, Eisenberg, and Simon 2012), but they do not model any product-level own or cross elasticities of advertising. Berndt et al. (1995, 1997) estimate the effect of marketing on both the size of the market and brand shares, focusing mostly on physician detail advertising and academic journal advertising because DTC was extremely limited and unbranded at the time, and found some effects at both category and product levels. In studying detailing effects, and to separate brand effects from category effects, Ching and Ishihara (2012) take advantage of the fact that different firms sometimes market identical molecules under different names in Canada. Narayanan, Desiraju, and Chintagunta (2004) estimate a two-level model using only time-series variation for antihistamines and do not find positive spillovers. In experimental work, Kravitz et al. (2005) find mixed results for patients going to their physicians asking for products they saw on television. In a structural model, Jayawardhana (2013) imposes that television advertising must affect only class-level demand and finds significant effects. Many of these studies either model only a category-level response or model only a conditional share-level response. This paper will model the full decision process and use data with both spatial and time-series variation. Liu and Gupta (2011) and Stremersch, Landsman, and Venkataraman (2013) also examine the various effects of DTC on aspects of demand.[1] Stremersch et al. can explain variation across geography using demographic characteristics and find heterogeneous effects. This study will differ from both of those studies in that I will use fixed effects to partial out the reasons for persistent differences in DTC across markets and focus on variation just across the borders.

The supply side of advertising in pharmaceuticals has been much less explored. If advertising helps rivals' demand, an incentive to invest less in advertising might well exist. Iizuka (2004) finds that as the number of com-

[1] Stremersch et al. (2013) look at the effects of DTC through the mediator of patient requests and find no effects. Liu and Gupta (2011) use information on patient visits.

petitors increases, firms advertise less, leading it to suggest the existence of a free-riding problem. Ellison and Ellison (2011) find evidence that pharmaceutical firms decrease advertising just prior to patent expiration in order to make the market smaller and deter generics from entering. The possibility of such strategic deterrence implies the existence of positive spillovers, at least from brand to generic. However, no research that I am aware of uses a supply model to quantify the magnitude of the potential positive spillover effects on advertising expenditure decisions. Ching (2010), Filson (2012), and Liu et al. (2016) use a Markov perfect equilibrium concept to model the supply side of pharmaceutical markets, but they focus on different aspects of the pharmaceutical industry rather than television advertising.

Outside of the pharmaceutical literature, Sahni (2016) finds experimental evidence of positive spillovers to rivals in online restaurant advertising in India. Additionally, Lewis and Nguyen (2012) and Anderson and Simester (2013) find evidence of positive spillovers in a number of categories for online and mail advertising, respectively. Nonexperimentally, Ching, Erdem, and Keane (2009) show evidence from scanner data that advertising of an individual brand with a display or feature could have spillover effects for the whole category.

This paper makes three contributions. First, I improve on the literature that seeks to identify the causal effect of advertising on own and rivals' demand with observational data by using an identification at the border approach. That is, I will identify advertising elasticities by comparing households that are very near to each other geographically but get different advertisements because of the way the television market borders are drawn. I show that advertising has significant positive effects on rivals' sales, though smaller than its effects on own firm sales. This is part of a growing literature seeking to identify the effect of advertising on demand, including the studies by Sinkinson and Starc (forthcoming), who use political advertising as an instrument, and Hartmann and Klapper (2017), who use the unexpected nature of the teams who play in the Super Bowl as a shifter of ad exposure. Second, I construct and estimate a consumer choice model that allows advertising to influence the size of the category, the conditional share of each subcategory in the category, and the conditional share of each product in a subcategory. I will consider the category, the subcategory, and the product levels as three separate stages of a joint physician-consumer decision-making process. At each stage, I will allow for advertising carryover effects. Results indicate that advertising of antidepressants affects both category demand and brand share. The category effects are larger and more persistent over time than are business-stealing effects, leading to a net positive spillover. Further, using the border strategy with fixed effects to identify the advertising parameters is important. Failing to use fixed effects to control for persistent differences in markets and systematic national changes over time in market conditions leads the researcher to conclude that advertising is primarily business stealing and drastically overstates the short- and long-run effectiveness. Failing to focus on the borders of television markets to control for the endogeneity of firm choices leads the researcher to overstate the long-run effectiveness of advertising and understate the relative long-run importance of category expansion relative to business stealing. Consistent with the free-riding incentives implied by the estimated parameters, I find that firms advertise less and less often in markets where rival advertising is high. Third, I conduct a supply-side analysis using a stylized model to evaluate the extent to which positive spillovers suppress the incentive to advertise. Given the demand parameters, I compute a benchmark competitive outcome in television advertising whereby firms optimally free ride. I find that if, instead, firms that advertise work together, removing the need for strategic response, those firms would combine to advertise 50 percent more than in competitive equilibrium. A cooperative deciding all advertising expenditure levels and taking full industry profits into account would advertise four times as much as is observed in competitive equilibrium and would increase the category size by 18 percent and category profits by 14 percent. These numbers are illustrative of the incentive effects of positive spillovers but should be seen as an upper bound on the magnitude of the true dynamic underinvestment due to positive spillovers, as the benchmark model assumes that firms optimally free ride. No other research that I am aware of conducts such a supply-side analysis of the provision of advertising that exhibits positive spillovers. This paper helps move us toward understanding the effects of advertising and the incentives facing the firms that provide it, and understanding both is essential to firm profit maximization and to efficient regulation.

II. Empirical Setting

A. Prescription Drugs and Advertising

Television advertising of prescription drugs did not appear in the United States until 1997. Although technically not forbidden by law, all advertisements were required to include much more risk information than is required today. This required risk information was similar to the package inserts that come with prescriptions. Reading those aloud in the context of a 30-second spot was prohibitively time-consuming and costly. In the fall of 1997, the FDA issued a draft memorandum clarifying its stance on advertising risk information, allowing advertisements to air as long as they had a "fair balance" of risk information, even if abbreviated. Firms had the opportunity to submit their advertisements to the FDA for preapproval to ensure they met the "fair balance" condition. In 1999, the final copy of the FDA memorandum was circulated. The first advertisements on television for antidepressants appeared in 1999 when GlaxoSmithKline's brand, Paxil, began airing its first campaigns.

Figure 1 suggests that the FDA regulation was binding prior to 1999, and advertising did not begin until that point.

1. Antidepressants

Prescription antidepressants are indicated for treatment of major depressive disorder and dysthymia, which is a more minor version of depression. Traditionally, depression was treated with what are called tricyclic antidepressants (TCAs), which were discovered in the 1950s but came with significant side effects and risks. Treatment of depression took a great leap forward in the late 1980s with the innovation of selective serotonin reuptake inhibitors (SSRIs), the first of which was Prozac. Newer-generation antidepressants are more tolerable than the older-generation TCAs and offer safer treatment and with fewer side effects (Anderson 2000). This breakthrough allowed easier management of antidepressant treatment by primary care physicians and made seeing a specialist less necessary.

Diagnosis and treatment of depression can be rather complicated, as with many mental disorders. As the class of drugs has grown, so has the number of people being treated. In 1996, the industry pulled in around $5 billion in revenue. By 2004, it was up to $13 billion. In 2004, an FDA black box warning was instituted suggesting antidepressants might lead

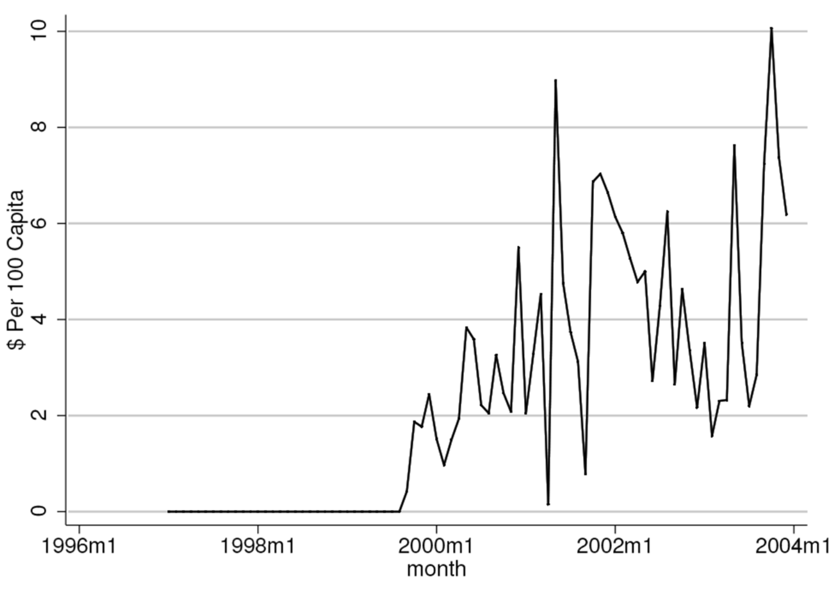

Fig. 1.—Antidepressant commercials relative to FDA memo

to an increase in suicidality among adolescents (Busch, Golberstein, and Meara 2014). Around the same time, many widely selling molecules began to go off patent. Figure 2 shows the revenues of the antidepressant industry from 1996 through 2004. Since the discovery of Prozac, 10 other brands, some with slightly different mechanisms, have been discovered and have entered the market. Some of those have developed extended-release versions that allow patients to have fewer doses per day.

Antidepressants have six main subcategories: the old style TCAs, tetracyclic, serotonin antagonist and reuptake inhibitors, serotonin-norepinephrine reuptake inhibitors, norepinephrine reuptake inhibitors, and SSRIs. Although the specific differences between these subcategories are not important to this study, note that each subcategory has somewhat different mechanisms, interactions, and side effect profiles from the others. Deciding which subcategory of antidepressant is appropriate for a given patient is largely up to the physician and often is related to other medications the patient is taking. The decision between drugs within a subcategory might depend on what the patient's insurance formulary or physician preferences include. Antidepressants are characterized by a high degree of experimentation to find a good fit between treatment and patient, as well as a low compliance rate due to the many side effects (Murphy, Cowan, and Sederer 2009).

Many physicians see depression as an undertreated condition, and some research has concluded that restricting access to antidepressants has been associated with negative health outcomes (Busch et al. 2014). Given this information, market-expansive advertising might play a role in this market.

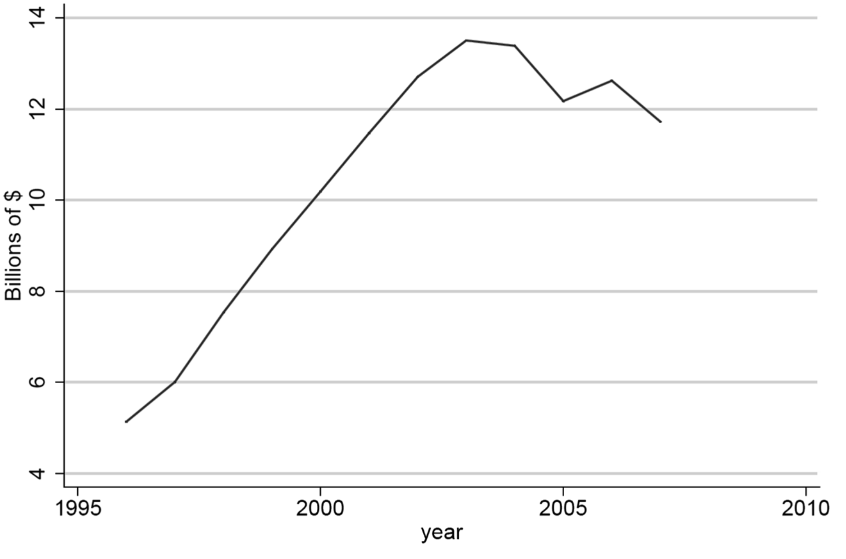

FIG. 2.—Antidepressant revenues, 1996–2008

2. The Market for Advertising

Firms can purchase television advertising space in two ways. First, each summer, in an up-front market, advertising agencies and firms make deals for the upcoming year of television. Advertising purchased in the up-front market cannot be "returned" and typically has minimal flexibility in terms of timing. Next, in a spot market that is called the "scatter" market, firms can purchase advertising closer to the date aired.

Additionally, both national and local advertisements are available for purchase. National advertisements are seen by everyone in the country tuned in to a particular station, whereas local advertisements are seen only by households within a particular designated market area (DMA).

A DMA is a collection of counties, typically centered around a major city, and is defined by AC Nielsen, a global marketing research firm. The DMAs were first defined to allow for the sale of advertising in a way that was straightforward to the advertisers. The DMA location of a county determines which local television stations a consumer of cable or satellite dish gets with his or her subscription. The original idea was to place counties into the same DMA with the local television station that most people wanted to watch, which often was the station that was easiest to pick up over the air. That is, if a county picks up the Cleveland stations over the air more easily than the Columbus stations, it would be placed in the Cleveland DMA. Existing laws and regulations in most circumstances do not allow satellite or cable operators to provide broadcast signals from outside of the DMA in which they reside.[2] Even for over-the-air signals, the Federal Communications Commission moderates the signals to try to keep the signal from each station localized only in its own DMA.[3] The United States has 210 DMAs, the largest 101 of which are included in my data.[4]

From informal conversations with individuals in the industry, I learned that pharmaceutical companies participate almost exclusively in the up-front market. As with most consumer goods, the majority of antidepressant spending is on national advertising, but there is a significant amount of local advertising as well as significant variation across DMAs in the amount of local advertising.

Projected volume and type of viewership typically determine prices for advertisements. A single airing of a national advertisement for antidepres-

[2] See http://www.sbca.com/dish-satellite/dma-tv.htm.

[3] See http://www.fcc.gov/encyclopedia/evolution-cable-television.

[4] Note that from time to time, Nielsen may move one county from one DMA into another. In these data, I have a snapshot of current DMA composition. Discussions with Nielsen have assured me these shifts are very infrequent. To the extent that a county gets categorized in the wrong DMA, it will lead to measurement error. Whether and how this miscategorization biases the estimation depend on the direction and magnitude of the correlation between the miscategorization and advertising. The direction and magnitude of any such correlations are untestable in the data.

sants ranges from $1,600 to $23,000 from 1999–2003, and a single airing of a local advertisement ranges from $0 to $7,600 for the same time period. Looking at each advertisement in terms of expenditure per capita, I observe that the distribution of local advertising expenditure per capita on a single commercial looks similar to the distribution of national advertising expenditure per capita on a single commercial. National advertisements range from $0.0002 per 100 to $0.04 per 100, and 93 percent of local advertisements fall within that range as well, with a few outliers going down to zero and up to $0.20 per 100 capita. By scaling expenditures by potential viewing population, local and national advertising expenditures are comparable.

Additionally, only four brands from three firms in this market advertise. Eli Lilly (Prozac, Prozac Weekly), Pfizer (Zoloft), and GlaxoSmithKline (Paxil, Paxil CR, Wellbutrin SR, Wellbutrin XL) are the only firms advertising in this market. Notably, those firms, along with Merck, are some of the largest advertisers within all of the pharmaceutical industry (Rosenthal et al. 2003). The lack of advertising from all firms could be indicative of fixed costs of advertising or of free riding. Those branded products that do not advertise either have low market share (Effexor XR, Remeron, Serzone) or have a very small parent company that might be less likely to have an advertising division (Celexa, Lexapro). Whether we can observe free riding will be further evaluated in the supply analysis.

B. Data

1. Prescribing Data

Sales data for this market come from the Xponent data set of IMS Health, a health care market research company. I follow the prescribing behavior of a 5 percent random sample of physicians who prescribe antidepressants, monthly from 1997 until 2004. The data include a rich set of physician characteristics including the address of the primary practice, which is then linked to a county. The data used in this study are aggregated to the county level and end in 2003, thereby avoiding confounding market changes in 2004, including the FDA black box warnings and wave of patent expirations. The sample is partially refreshed annually.

2. Advertising Data

Product-level monthly advertising data at the national and DMA levels for the top 101 DMAs come from Kantar Media. In addition to advertising expenditures, the data include the number of commercials. The unit of advertising used in this study will be expenditures per 100 capita in the viewing area. Scaling expenditures by population in the viewing area allows me to have a comparable measure of advertising volume between national

TABLE 1
Descriptive Statistics: Advertising

	Mean	Quarter 25	Median	Quarter 75
DTC per 100 capita	.782	0	0	1.358
Subcategory DTC per 100 capita	2.012	0	1.496	3.505
Category DTC per 100 capita	4.035	2.284	3.515	5.534
DMAs	101			
DMA population	2,340,774	903,090	1,469,823	2,622,567

and local advertising. Total advertising for a county is defined as the national advertising expenditure scaled by the national population plus the local advertising expenditure scaled by the population of the DMA.[5] Table 1 provides descriptive statistics for the DMA-level advertising variables at the product, subcategory, and category levels for the period of the data during which advertising is allowed: September 1999 through December 2003. The statistics are also only on the products that ever advertise: Paxil, Paxil CR, Prozac, Prozac Weekly, Wellbutrin SR, Wellbutrin XL, and Zoloft.

Figure 3 depicts local advertising expenditures per 100 capita in Boston, New York, and Austin, Texas, as well as national advertising as examples of what local advertising expenditures look like over time. Local advertising for Paxil is higher in New York than it is in Boston, which in turn is higher than it is in Austin, suggesting that nontrivial variation occurs across markets in this measure. National advertising makes up the bulk of the advertising that households see, but the local additions to the national advertising vary considerably.

3. Detailing Data

In addition to DTC data, I have collected physician-level detailing data from ImpactRx, a market research firm. In the data, a panel of 2,134 general practice physicians are followed monthly from 2001 through the end of 2003, and a panel of 167 psychiatrists are followed monthly from 2002 to 2003. This panel is a national and geographically representative sample of physicians, most of whom are in the 40th percentile or greater in terms of total prescriptions written. This nonrepresentativeness arises because these physicians are the most likely ever to be detailed. Although these physicians make up less than 1 percent of total physicians in the country, they are likely to make up a significantly higher percentage of both the prescription and

[5] A possible alternative measure would be the number of commercials at the national level plus the number of commercials at the local level. I explored using that measure, and the results were not qualitatively different. However, because a commercial during the evening news is likely to capture far more eyeballs than a commercial during a 1:00 a.m. rerun of *MacGyver*, using expenditures per 100 capita would seem to do a better job at measuring quality-adjusted advertising than number of commercials.

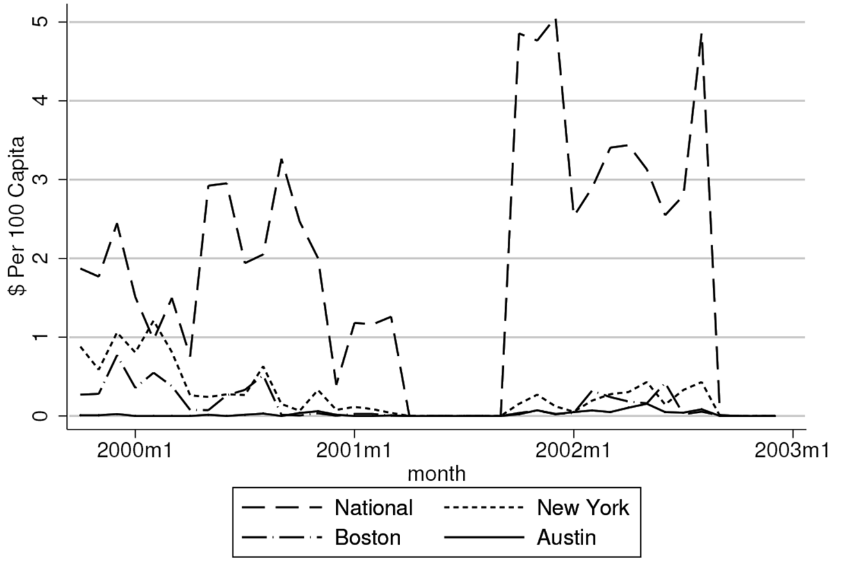

FIG. 3.—Variation across three markets in advertising

detailing distributions. Additionally, national aggregate detailing data by brand from IMS Health are observed in the data.

4. Other Data Sources

I observe prices from Medicaid reimbursement data, collected by the Centers for Medicare and Medicaid Services (CMS). Duggan and Scott-Morton (2006) argue that the average price Medicaid pays per prescription prior to Medicaid rebates is a good measure of the average price of a drug on the market. As my measure of price, I use the total Medicaid units dispensed divided by the total Medicaid reimbursements during a quarter for a particular product, deflated to 2010 dollars using the consumer price index.

CMS also collects data on the average pharmacy acquisition cost for all pharmaceutical products (NADAC). Because I will not be estimating marginal production costs empirically, I can use these average pharmacy acquisition costs as an effective upper bound on marginal production costs. Although markups from branded drugs exist, pharmacies are typically able to obtain generics at much lower rates, particularly when several generic competitors are present (as is the case in this market), often as low as 10 cents per pill. As of 2013, all products in the sample have generic versions available. For an upper bound on the marginal cost of each drug, I use average pharmacy acquisition cost for those generic versions of the product, deflated to 2010 dollars using the consumer price index.

Yearly county population, employment, demographic, and income data are drawn from the Current Population Survey.

III. Reduced-Form Evidence

In this section, I explore the data to see if spillover effects exist and how they interact with own effects. Estimates show that rivals' and own advertising have a positive effect on sales, whereas rivals' advertising has a smaller effect than own advertising. In addition, the cross partials indicate that rivals' advertising makes own advertising less effective, but own advertising has a larger negative effect on the marginal own advertisement because of decreasing returns to scale.

In particular, I model sales of quantities $\boldsymbol{Q}$ of product j in time t for market m as a function of own advertising, a^{own}, and advertising of rivals, a^{cross}:

$$\begin{aligned}\log(Q_{jmt}) = {} & \lambda\log(Q_{jm,t-1}) + \gamma_1 a_{jmt}^{\text{own}} + \gamma_2 a_{jmt}^{\text{cross}} + \gamma_3(a_{jmt}^{\text{own}})^2 + \gamma_4(a_{jmt}^{\text{cross}})^2 \\ & + \gamma_5 a_{jmt}^{\text{own}} a_{jmt}^{\text{cross}} + \varepsilon_{jmt}.\end{aligned} \tag{1}$$

This estimation provides insight on whether rivals' advertisements help or hurt own demand, the nature of decreasing returns to scale, and persistence in advertising effects.

A. *Empirical Identification Strategy: Border Strategy*

The endogeneity of advertising and the absence of obvious instruments pose challenges to causal identification of the effect of advertising on demand.

I address the endogeneity concerns associated with advertising decisions by taking advantage of the discrete nature of local advertising markets. That is, two households that are directly across the television market border from one another will see different advertisements despite being otherwise very similar households. I take advantage of this comparison. This approach is similar in spirit to that used by Card and Krueger (1994) and Dube, Lester, and Reich (2010) to identify the effects of minimum wage increases and that used by Holmes (1998) to identify the effect of right-to-work laws. These three studies rely on state borders, across which any number of laws, market conditions, or preferences may vary. Similar spatial strategies have also been used by Black (1999) and Bayer, Ferreira, and McMillan (2007) using school zone borders and Ito (2014) using electricity market borders. A nice feature of television market borders is that they were set with television in mind and have little correspondence with anything else in the world. As such, we might think that the location of DMA borders is far more exogenous to consumer characteristics than are state borders.

Advertising is purchased both nationally and locally. The DMA to which the household's county belongs, as defined by AC Nielsen, determines the level of total advertising aired in a household's market. Nielsen places counties into markets by predicting the local stations in which the households will be most interested. As such, DMAs tend to be centered in metropolitan areas. A map of all of the DMAs included in the advertising data is presented in figure 4.

To get an idea of how advertising is distributed across the country, consider the example of the Cleveland and Columbus DMAs. Figure 5 depicts the state of Ohio with each DMA in a different color. Every county in the Cleveland DMA gets the same amount of the same advertising as every other county in the Cleveland DMA. Meanwhile, every county in the Columbus DMA gets the same amount of the same advertising as every other county in the Columbus DMA, though this amount might be different from the amount of advertising in the Cleveland DMA. Meanwhile, these two DMAs border each other. Five counties in the Cleveland DMA share a border with at least one county in the Columbus DMA, and five counties in the Columbus DMA share at least one border with a county in the Cleveland DMA. My strategy will be to consider these 10 counties as an experiment with two treatment groups (Cleveland and Columbus) in each time period.

The data contain 153 such borders. The map of all of the counties included in this border sample is presented in figure 6. Each of these borders will be considered a separate experiment, with the magnitude of the treatment determined by the advertising in each DMA at a given time. Only the counties bordering each other will serve as controls for each other to partial out any local effects that may be increasing or decreasing for both sides of the border. The level of an observation is a product-border-DMA-month, which means that a group of counties along a particular border but in the same DMA are aggregated together, because they each see the same advertising and they are each being compared with a similar group across the border. In each "experiment," one such set of counties will be compared with an adjacent set of counties across the DMA border. Each border experiment in each time period will include two observations: one for the group of counties on one side of the border and one for the group of counties on the other side of the border. Each of these observation groups will constitute a market.

To estimate the effects of advertising in this experiment, I will use a modified difference-in-differences estimator. The identifying assumption is that along the border of two DMAs, any differential trends in demand between the two sides of the DMA border stem from differences in advertising. In particular, I use panel data with fixed effects. Border-time fixed effects will ensure that the common trend assumption is enforced locally only at the border between two DMAs, allowing for spatial heterogeneity. Border-DMA fixed effects will allow systematically different demand levels across

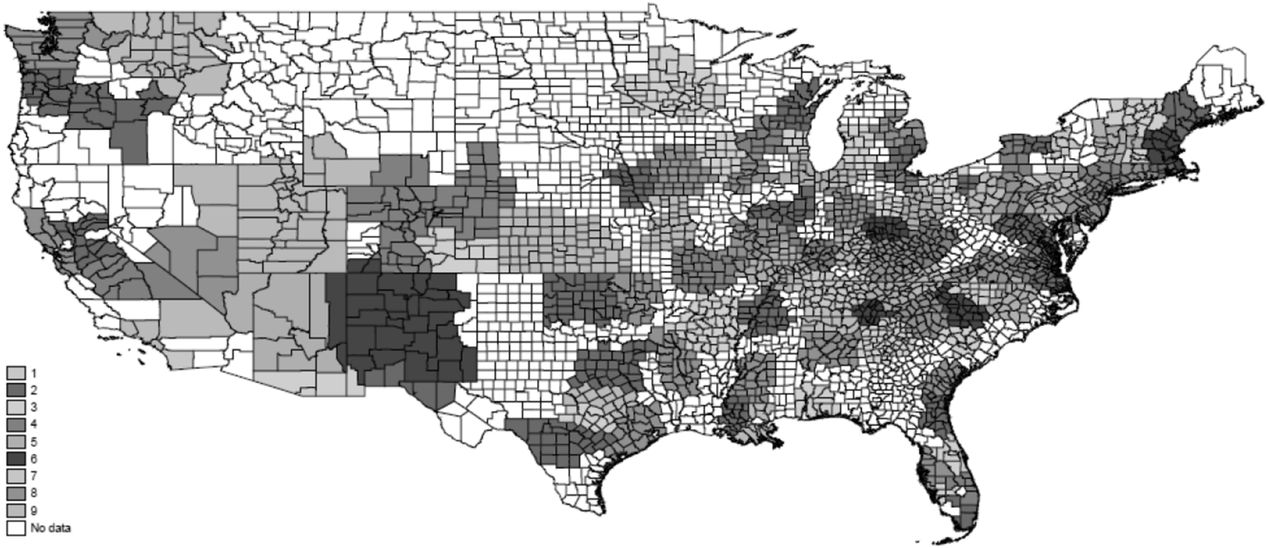

Fig. 4.—Full sample: top 101 DMAs

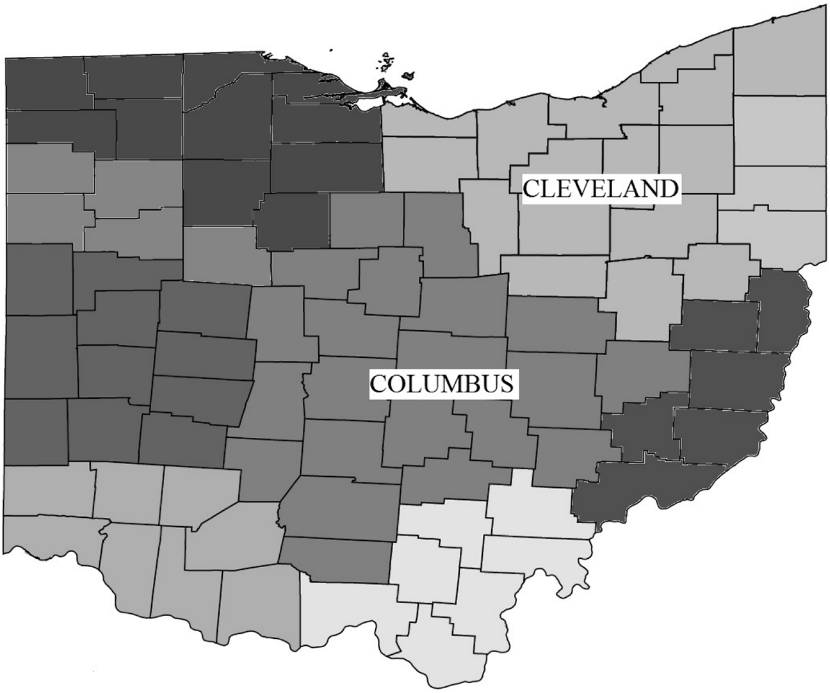

FIG. 5.—Ohio and its DMAs

the border. I will also include a lagged dependent variable to get at the dynamic effects of advertising. Consider the log of quantity $\log(Q_{jbmt})$ at the product-border-DMA-month level. Advertising, a_{jmt}, as mentioned before lives at the product-DMA-month level and affects $\log(Q_{jbmt})$ through some function f:

$$\log(Q_{jbmt}) = f(a_{jmt}) + \varepsilon_{jbmt}.$$

Each product-border pair will constitute an experiment with border markets as treatment groups. The fixed-effects specification is

$$\log(Q_{jbmt}) = \lambda\log(Q_{jbm,t-1}) + g(a_{jmt}) + \alpha_{jbq} + \alpha_{jbm} + \varepsilon_{jbmt},$$

where the subscripts j and b indicate which experiment is being considered (product- and border-specific); α_{jbq} is a time effect that is used to control the experiment, which in this case will be a quarter fixed effect; α_{jbm} is a treatment group fixed effect; and $g(a_{jmt})$ is the magnitude of the treatment. The magnitude of the treatment is zero everywhere prior to 1999, because the FDA memo had not yet gone into effect. To investigate persistence in demand, a lagged dependent variable is also included. Note that the inclusion of α_{jbm} in the specification means I am focusing on market-level

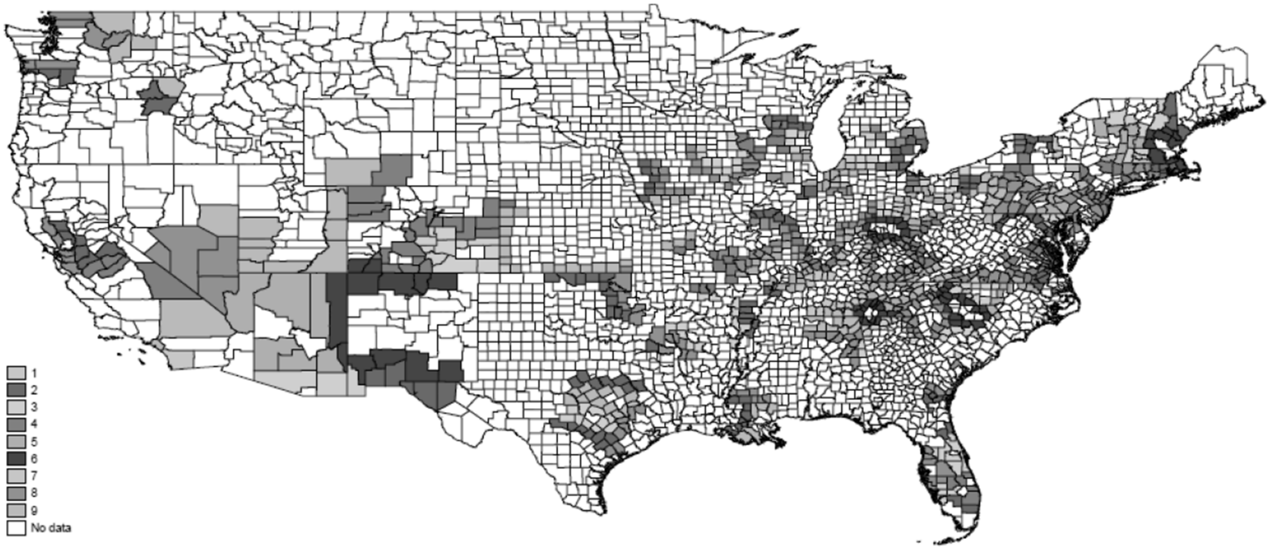

Fig. 6.—Border sample: counties on the borders of the top 101 DMAs

deviations from trend. That is, each market has a fixed effect. Whereas Stremersch et al. (2013) find that region-specific demographic composition can explain the distribution of DTC advertising across markets, such cross-market-level variation in advertising is accounted for in this specification with the fixed effect. The remaining variation being used is within-market, within-quarter deviations from the border experiment-specific common time effect.

For further intuition, again consider the Cleveland-Columbus example and the case of Zoloft advertisements. In the equation above, $\log(Q_{jbmt})$ is the log number of prescriptions of Zoloft in the Cleveland-Columbus border, indexed by month and which side of the border it is on. The magnitude of the treatment, $g(a_{jmt})$, is a function of Zoloft's advertising in each market. The time effect, α_{jbq}, is a common quarter fixed effect between the Cleveland and Columbus sides of this border and is used to subtract out contemporaneous macro effects. The fixed effect, α_{jbm}, allows the different sides of the border to have systematically different levels in the outcome.

For this strategy to be valid, the Cleveland and Columbus sides of the border may differ by a fixed level, but they must have common trends in the absence of advertising differences. Is this assumption plausible? These counties are bordering, so they are similar in geography. Both are sufficiently far from their central cities. The counties on the Cleveland side are only slightly closer to Cleveland than they are to Columbus and vice versa.

Also worth noting is that if Columbus always had a high, constant level of advertising and Cleveland always had a low, constant level of advertising, this estimation strategy would have no power to identify the effects of interest, because the border-DMA fixed effect would subtract out this variation, even though that advertising in Columbus might well have had an effect. In the sample period, at least some variation will be present in each experiment over time.

Although advertising is clearly a firm choice rather than completely random, thinking about the potential sources of endogeneity and how the border strategy addresses those specific sources is instructive. Because market-level fixed effects exist, endogeneity that comes from, say, the fact that winters in Florida are milder than winters in Wisconsin is not a concern. The market-level fixed effects absorb those types of concerns. Potential bias can come only from within-market, time-specific demand shocks that affect the firm choice of advertising. Those shocks could come from two main sources: unobserved events (unseasonably bad weather, a large local employer laying off a large number of workers, an important medical seminar that discusses the virtue of these drugs) or rule-of-thumb-based decision making.

First, consider the possibility of unobserved events. Because firm advertising decisions are made at the DMA level, the unobserved shocks of interest are the average across the DMA. Consider an unseasonably cold month that makes people more depressed, boosting both advertising and prescriptions of antidepressants. Weather patterns are continuous phenomena in that the weather should not be significantly different on one side of a county border versus another. However, over larger distances, weather tends to be very different. As such, the average temperature over the DMA might be much colder than it is at the border of the DMA, but at the border, the temperature will be similar on both sides. The border strategy takes care of this type of endogeneity. Similarly, consider a large shock to employment in a given month in a DMA. This shock might simultaneously lead to a large increase in depression as well as an increase in advertisements, potentially biasing any estimated effect of advertising. Employment tends to be more concentrated in cities, which tend to be at the center of DMAs. The farther away a person is from a place of employment, the less likely he or she will work there, because of costs of commuting long distances. The distances just across the DMA borders to a central city are pretty similar and do not discontinuously jump as the border is crossed. As such, at the border, counties bordering but on opposite sides of the border are similar in their potential to be affected by any particular employment shock, but they are much less likely to be affected than those close to the center of the DMA. Again, the border strategy should be able to handle this source of endogeneity. Finally, consider a seminar meant to educate physicians about any particular course of treatment. This seminar might increase the use of some antidepressant while also increasing advertising to the DMA where it occurs. Because these seminars also tend to be in centers of DMAs, transportation costs are likely to prevent those physicians at the outskirts from attending, but those transportation costs do not discontinuously change at the DMA border. These three unobserved shocks seem to be the most likely ones to drive advertising decisions at the DMA level, and all are addressed by the border approach. Note that other potential unobserved shocks might exist. However, as long as the strength of these shocks diminishes reasonably continuously in the distance from the center of the DMA and does not discontinuously change at the border, the border approach will be valid. All these requirements hold for the aforementioned examples.[6]

Next, consider the possibility that firms use rules of thumb to allocate advertising on the basis of the demand in the previous period, in which case, advertising in a DMA in the current period is determined by some

[6] Although most DMAs are centered around a large city, a few DMAs have two main cities (e.g., Johnstown-Altoona, PA). If the main targets of a DMA are at the borders, for the border strategy to remain valid, only the demand shocks immediately across the border need to be the same, because the primary worrisome unobserved demand shocks described above are spatial in nature. If the firm targets those specific demand shocks, variation is unlikely to exist across the border, because the firm will want to target both DMAs. In that case, the particular experiment is not helpful to identification but also not particularly harmful. If the firm targets those demand shocks in only one of the two DMAs, the variation will still be valid for identification, because we are controlling for the demand shocks directly using the borders.

function of last period's demand across the DMA. If previous-period demand is correlated with current-period demand, estimates will be biased. This scenario is a classic reverse causality issue in advertising noted by Berndt (1991) and Bagwell (2007), among others. How does the border sample help this problem? The border counties make up a fraction of the population, sales, and counties in a DMA. Even if the trends for demand are identical between the border areas and the DMA as a whole, the covariance between last period's demand in a DMA and current demand in a border area is a small fraction of the covariance between last period's demand in the whole DMA and this period's demand in the whole DMA. Further, the demand trends are likely to be different between the border and the full DMA, further reducing that covariance and reducing any omitted variable bias. Again, by comparing the counties along the border to their counterparts on the other side of the border via the common time trend, the omitted variable of last period's demand will be absorbed into the product-border-time fixed effect, because demand is similar immediately across the DMA borders.

In principle, the rule-of-thumb reverse causality problem could be solved using the full DMA but controlling for the previous period's demand. That approach is problematic in this setting because the previous-period demand is already in the model and has an interpretation of its own. If previous-period demand is effective as a control for rule-of-thumb advertising, it will inflate the estimate on the lagged dependent variable, and its interpretation as a persistence parameter will be incorrect. The border approach, by making comparisons between similar counties that constitute a small fraction of total DMA demand, alleviates the concern of reverse causality.

1. Limitation of the Border Strategy

The main limitation of the border strategy is similar to that of a usual regression discontinuity design in that the estimated treatment effects are identified at the border and not elsewhere. The true treatment effect in the interior of the DMA might be different from that at the border, in which case, the interpretation of the supply analysis will be limited because I assume that the estimated advertising effects hold in both border and nonborder counties. A robustness check in the online appendix partially addresses this concern. In particular, I separate out borders that are closer to urban centers and estimate the model separately for those borders than from those that are farther out. The effects are similar and not statistically distinguishable from those of the full border sample. Although this result might not fully establish that the effect in the interior of the DMAs is the same as the effect at the borders, it provides a small piece of evidence that the effects seem to hold up similarly across different types of counties. Further, I compare measurable characteristics for in-sample counties (those at the borders) with the characteristics out of the sample (those at the interior). Those comparisons are available in appendix F.2. A t-test fails to reject that border counties and interior counties are the same in average population, average income, average number of physicians, and number of nonfederal physicians. To the extent that we continue to worry that the estimated treatment effects are different at the interiors, it will be a limitation of this analysis. Similar limitations apply to other regression discontinuity designs as well as to instrumental variables methods that reveal only local average treatment effects.

2. Potential Threats to the Border Strategy

One potential worry is that little variation would exist net of the fixed effects, which would be the case if too much of the advertising was national and not enough was local. Figure 7 displays a histogram of advertising net of these fixed effects showing significant variation. Net of fixed effects, the log of advertising expenditures per 100 capita has a mean of zero and a standard deviation of 0.25, so substantial variation exists net of fixed effects.

Also potentially problematic is the lagged dependent variable, which can generate omitted variable bias in the presence of small T, because differencing mechanically induces correlation between the lagged depen-

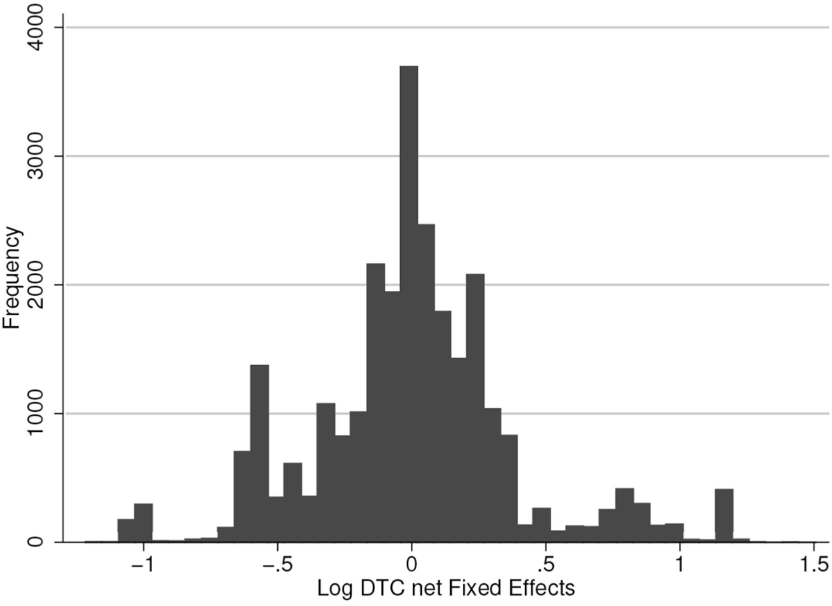

FIG. 7.—Variation in log DTC net of fixed effects

dent variable and the error term. However, as $T \to \infty$, the mechanical correlation with the error term diminishes to zero and the fixed-effects estimator is consistent. Because my data are monthly from 1997 through 2003, $T = 84$ should be sufficiently large that any bias will be minimal.[7]

Additionally, we might be concerned about measurement error. Three main possibilities could lead to measurement error and biased estimates:

1. Consumers watch advertisements in one DMA but drive across the border to see their physicians.
2. Consumers watch advertisements in one DMA but drive toward the center of the DMA to a county not included in the border sample to see their physicians.
3. Consumers watch advertisements over the air and sometimes see the advertisements from the DMA that is on the other side of the border.

All of these scenarios would lead this approach to understate the effect of the advertisements. To the extent that we think these biases are present, we view these estimates as lower bounds on the true parameters.[8]

Also note that few consumers watch over the air. According to the Consumer Electronics Association, fewer than 7 percent of households rely on over-the-air signals for their television.[9] Further, at the DMA border, TV signals tend to be less reliable over the air, because stations tend not to locate in the outskirts of DMAs. Consumers would probably be even less likely to rely on over-the-air signals at the DMA border.

I should also note that there could be other sources of omitted variable bias. I omit prices, magazine and newspaper advertising, and detailing from

[7] In Nickell's (1981) paper, which describes the bias induced when doing fixed effects and lagged dependent variables, he analytically solves for the bias as a function of T. The bias is

$$p\lim_{N\to\infty}(\hat{\lambda}-\lambda) = \left\{\frac{2\lambda}{1-\lambda^2} - \left[\frac{1+\lambda}{T-1}\left(1-\frac{1}{T}\frac{1-\lambda^T}{1-\lambda}\right)\right]^{-1}\right\}^{-1} \approx \frac{-(1+\lambda)}{T-1},$$

where the approximation holds for "reasonably large" T. With $T = 84$, that approximation is bounded above by $-2/83 \approx -0.02$. If we do not wish to concede that 84 is reasonably large, plugging in 0.7 as the true λ, the exact bias formula gives the bias at -0.025. Because this holds only as $N \to \infty$, I ran simulations assuming that the data-generating process is as estimated and found the magnitudes to be nearly identical. Details of the simulation are available from the author on request.

[8] However, the Dartmouth Institute has drawn primary care commuting zones that describe how far Medicare patients travel to see their physicians. Commuting zones rarely cross DMA lines: only about 1 percent of primary care commuting zones cross DMA borders, and those that do tend to be predominantly in only one DMA. This fact should minimize the measurement error worry. Further explanation of the Dartmouth Institute commuting zones is provided in the appendix.

[9] See http://www.tvtechnology.com/default.aspx?tabid=204&entryid=9940.

this estimation. Because prices tend not to vary geographically because of a very low transport cost and ease of obtaining drugs through the mail, prices are absorbed in the product-border-time fixed effect. Similarly, magazine and newspaper advertising is all in national publications for this category. These national publications might have differential take-up across geography. Because I will be employing time fixed effects at the border level, the take-up of these national publications must follow parallel trends across the DMA borders for them not to contaminate the results here. Because this study does not have data on differential penetration rates of national publications at the county level, parallel trends in these variables will remain an assumption. The product-border-time effect also controls for any national average effects of detailing.

However, firms strategically raising (or lowering) detailing at the product-market-time level in exactly the same places where DTC is concentrated to take advantage of any complementarities or substitutabilities will induce bias. For a monthly panel of 2,134 general practitioners and 167 psychiatrists in 2001–3, ImpactRx data have physician-specific detailing information on the number of sales representative detail visits. In appendix A, I show that for any given time period and market, DMA totals of detailing visits are uncorrelated with DTC. I show this absence of correlation by aggregating physician-level monthly detailing visits to the DMA level and running a regression with number of detailing visits on the left-hand side and own and rival DTC advertising on the right-hand side, as well as product-DMA and product-time fixed effects. The coefficients on own and rival DTC are small and insignificant. I further show that this result extends to the border areas, by aggregating physician-level detailing visits to the border-experiment level and running the same regression including product-border-DMA and product-border-time fixed effects. Again, the coefficients on own and rival DTC are small and insignificant. Omitting detailing from the main model specification puts any detailing effects on demand into the error term. As long as this error term is orthogonal to DTC advertising, the omitted variable bias will be zero. The above provides evidence that the detailing component of the error term is in fact orthogonal to DTC. I confirm this intuition by including border-experiment-level detailing visits into the main model in the paper for only the dates 2001–3 and markets for which detailing data are available and show that the inclusion of detailing does not affect estimates of the effect of DTC. Because the number of time periods is greatly reduced when using these detailing data, my preferred specifications will omit detailing and use the long time series.

A further piece of evidence against the coordination of DTC and detailing comes from the IMS national aggregate detailing data. In September of 1999, the FDA introduced guidance making DTC advertising feasible in the United States. If detailing is significantly coordinated with DTC, we would expect to see a discontinuous change in detailing when the law change causes

DTC advertising to increase significantly. In appendix A, I show that no trend break occurs in firms' detailing strategy nationally at that point at which DTC changes drastically from zero to significantly positive.

This lack of coordination may seem surprising, because detailing and DTC advertising are two important pieces of the marketing mix for pharmaceutical firms. Firms have strong institutional reasons for possibly failing to coordinate these efforts during the sample period. In particular, the organization of the firm and the nature of sales representative employment make coordination difficult. Those managers who decide DTC advertising tend to be in the consumer division of the firm, whereas those who coordinate detailing are in the sales division of the firm. Further complicating the coordination problem, sales representatives are generally independent contractors. That is, the firm, through a typically external analytics company, makes suggestions of how many times each sales rep should visit each physician. These suggestions are largely based on decile rules. The literature documents these decile rules (Manchanda, Rossi, and Chintagunta 2004). The basic intuition is that the analytics company groups physicians into 10-decile buckets based on prescription volume. They then suggest that sales representatives visit each physician in the same decile the same number of times. The deciles are adjusted over time as physicians change their intensity of prescribing, but adjustments more frequent than yearly tend to be minor.[10] Each sales rep then has the option to follow or not follow those recommendations and is compensated for the eventual prescriptions written by the physicians visited. For systematic coordination of DTC with detailing to exist, either the sales reps would all have to decide to work harder during high-DTC months or more sales reps would have to be temporarily hired. Neither of these possibilities is easily executed, particularly on a month-to-month basis.

A further concern is that various policies or cost inputs could discontinuously change at the DMA border, causing the more impressionable physicians to locate on a particular side of the DMA border. Because DMAs are in general relevant only to television markets, imagining why any tax laws would systematically vary across DMA borders is hard. Almost all business tax policies are set by state governments or potentially large city governments. Being on the border of the DMA typically leaves those counties out of reach of large city-specific taxes. However, because many DMA borders coincide with state borders, state tax policies could be a problem. To address that concern, I have removed the DMA borders that coincide with state borders and reestimated the whole model. The results of the estimation are available in appendix C. The estimated parameters are not statistically different from those if all borders are left in the sample.

[10] Conversations with experienced sales representatives confirm this intuition.

Data on corporate rental rates by county are not available for this study, so assessing whether a discontinuity is present in the rental rates of physician offices across a DMA border is impossible. However, imagining rental rates would differ significantly within a state, for two very similar counties that border each other and are similar distances from major cities, is hard. Furthermore, if they did differ, imagining that advertising decisions for the full DMA would hinge on rental rates at the border of the DMAs is hard. The only worry is if the physicians who select into cheaper rent are systematically those who have different responsiveness to advertising than those who select into more expensive rent. To further examine the issue of selection across the border, I collected data from the Area Resource File to see if the number of physicians, the average income, or the population on the higher advertising side of the border differs significantly from the population on the lower advertising side of the border. *T*-tests cannot reject that all of these variables are the same across the borders. Results of these tests of balance are available in appendix E.

Finally, the identifying assumption of difference-in-differences could be violated. The difference-in-differences model might fail the parallel trends assumption, invalidating the difference-in-differences design. To address this concern, I have conducted a placebo test. Using data on DMA-level television advertising of over-the-counter sleep aids as a placebo treatment, I find no economically significant effects. Details for this robustness check are in the appendix.

3. Why the Border Strategy?

A more conventional identification strategy in the discrete-choice literature is to use an instrumental variables approach, as in Berry, Levinsohn, and Pakes (1995). The main identifying assumption for the validity of the Berry et al. instruments is that the characteristics of competing products within a market are exogenous; thus, the changing competitive structure of the market may be used as a supply-side instrument for demand-side choice variables. In the market for prescription drugs, entry happens in all markets simultaneously by all products; thus, the use of the Berry et al. instruments would eliminate any spatial variation, which is a main attribute of the data I am using. Furthermore, thinking that competitor characteristics are exogenous in this setting might be unreasonable. It stands to reason that as consumers demand more antidepressants with fewer side effects, firms might well focus research and development on that kind of product.

B. Results

Using the identification strategy at the border outlined, the estimating equation including fixed effects becomes

$$\log(Q_{jmt}) = \lambda \log(Q_{jm,t-1}) + \gamma_1 a^{\text{own}}_{jmt} + \gamma_2 a^{\text{cross}}_{jmt} + \gamma_3 \left(a^{\text{own}}_{jmt}\right)^2 + \gamma_4 \left(a^{\text{cross}}_{jmt}\right)^2 + \gamma_5 a^{\text{own}}_{jmt} a^{\text{cross}}_{jmt} + \alpha_{jbq} + \alpha_{jbd} + \varepsilon_{jmt}, \tag{2}$$

where α_{jbq} is a product-border-quarter fixed effect and α_{jbd} is a product-border-DMA fixed effect. The α_{jbq} effect will sweep out all variation that is not between two areas that are on opposite sides of a DMA border. The product-border-DMA fixed effects sweep out all variation that is due to persistent differences between different markets (e.g., people are generally more depressed in New York than in Wisconsin).

Partialing out these fixed effects makes the identifying variation within product j local advertising that is over and above the average on its side d of the border b and over and above the average local advertising of product j in time period t in all counties on either side of border b.

Results of the above regression are provided in table 2. Most notable is that both rivals' and own advertising have a positive and significant effect on demand. Rivals' advertising hits decreasing returns to scale more slowly than does own advertising. Also, the cross partial indicates that rivals' advertising works a firm down its marginal revenue curve with respect to advertising, but not as much as own advertising does. This negative effect of the cross partial is the source of the incentive to free ride, as rival advertising lowers the marginal revenue of own advertising. Finally, evidence of persistence is present, though the persistence parameter is not especially

TABLE 2
The Effect of Own and Rival Advertisements on Sales

Variable	Log(Q)
Lagged log(Q)	.334***
	(.00746)
DTC	.0240***
	(.00621)
DTC^2	−.00216*
	(.00113)
DTC$_{\text{rival}}$	.0164***
	(.00266)
DTC$^2_{\text{rival}}$	−.000938***
	(.000252)
DTC × DTC$_{\text{rival}}$	−.00134**
	(.000631)
Product-border-time	Yes
Product-border-DMA	Yes
Observations	316,428
R^2	.955

Note.—Product-DMA clustered standard errors are in parentheses.
* $p < .1$.
** $p < .05$.
*** $p < .01$.

large. This finding is consistent with the idea that finding the correct fit between patient and treatment in the depression space requires experimentation.

IV. Model

A. Demand

I propose a multistage choice model in which advertising may affect the consumer's choice at each stage. A consumer arrives at her desired end product through a sequence of choice problems. First, she chooses between entering the category (inside option) and the outside option. If she chooses to enter the category, she chooses which subcategory of product she wants. Finally, given her choice of subcategory, she chooses which product to purchase. This process can be extended, in principle, to have any number of stages.

In the specific case of prescription antidepressants, this scenario is plausible. A consumer first decides whether she has a problem with depression, goes to the physician, and, together with the physician, determines which class of drugs would be most suitable (perhaps considering interactions with other drugs taken) and which product in particular is the best choice (perhaps having to do with what is on her formulary). The basic structure of this demand model is similar to those in Berndt et al. (1997) and Ching et al. (2016). The approach is slightly different from that in Liu, Steenburgh, and Gupta (2015), which includes rival product characteristics directly in the own utility function to account for possible spillovers.

I define "utility" u of consuming the inside option to be a function of total advertising stock as well as other market-level factors:

$$u_{ilmt} = \Gamma_1(A_{lmt}) + \beta_1 X_{lmt} + \alpha_{lt} + \alpha_{lm} + \xi_{lmt} + \varepsilon_{ilmt} = \delta_l + \varepsilon_{ilmt}. \quad (3)$$

In this specification, l denotes the inside versus outside option, m denotes market, and t denotes time period. I define Γ_1 as an increasing function of A_{lmt}, total advertising stock of all inside option products in market m at time t, α_{lt} as a time-specific taste for the inside option, α_{lm} as a market-specific taste for the inside option, and X_{lmt} as market-time characteristics.

For the next stage, I define the relative utility v^l of subcategory n conditional on the choice of the inside option as a function of the total advertising stock in subcategory n, A_{nmt}, as well as other subcategory market-time-level factors:

$$v^l_{inmt} = \Gamma_2(A_{nmt}) + \beta_2 X_{nmt} + \alpha_{nt} + \alpha_{nm} + \xi_{nmt} + \varepsilon_{inmt} = \delta_{n|l} + \varepsilon_{inmt}. \quad (4)$$

Finally, relative utility w^n of product j conditional on the choice of subcategory n is defined as a function of advertising stock of product j, A_{jmt}, and other product-market-time-level factors:

$$w^n_{ijmt} = \Gamma_3(A_{jmt}) + \beta_3 X_{jmt} + \alpha_{jt} + \alpha_{jm} + \xi_{jmt} + \varepsilon_{ijmt} = \delta_{j|n} + \varepsilon_{ijmt}. \tag{5}$$

Dynamics enter the model through advertising carryover. That is, a consumer may remember an advertisement from a previous period, and that advertisement may affect current-period demand. In general, advertising stock is a function of current-period advertising (measured in expenditure per 100 capita) in choice stage s, a_s, where $s \in \{l, n, j\}$; last period's advertising stock, $A_{sm,t-1}$; and a parameter governing depreciation over time, λ_s:

$$A_{smt} = f(\lambda_s, A_{sm,t-1}, a_{smt}). \tag{6}$$

I set each disturbance term, ε, to be independent and identically distributed extreme value type I. Given the logit errors, I compute a closed-form solution for shares. The unconditional share of product j in subcategory n is a product of conditional shares, where market and time subscripts have been suppressed:

$$s_j = (s_{j|n})(s_{n|l})(s_l). \tag{7}$$

Those conditional shares take logit form

$$s_{j|n} = \frac{\exp(\delta_{j|n})}{1 + \sum_{j \in n} \exp(\delta_{j|n})}, \tag{8}$$

$$s_{n|l} = \frac{\exp(\delta_{n|l})}{1 + \sum_{n} \exp(\delta_{n|l})}, \tag{9}$$

$$s_l = \frac{\exp(\delta_l)}{1 + \exp(\delta_l)}. \tag{10}$$

I note here that the error terms at each level are independent of each other.[11] I allow each level to have a different persistence, λ_s, and different effects of advertising, Γ_s. I also note that although I call the latent variables at each level "utilities," interpreting them literally as such is not es-

[11] Because each equation is a "conditional" statement, the independence of the error terms seems reasonable. That is, the error term at the business-stealing level indicates, conditional on already having chosen to get an antidepressant and having decided that an SSRI is appropriate, what is my idiosyncratic taste for Prozac versus Zoloft? Imagining why a relative preference between Prozac and Zoloft should affect a consumer's absolute taste for antidepressants is hard.

TABLE 3
Demographic Descriptive Statistics: Border Sample, 1997–2003

Variable	Mean	Standard Deviation	Minimum	Maximum
Percent black	.0744	.0994	0	.628
Percent Hispanic	.0647	.102	.00209	.774
Percent Asian	.0137	.0245	.00034	.290
Percent urban	.596	.253	0	.999
Percent uninsured	.154	.046	.076	.334
Percent over 45	.378	.0501	.217	.571
Percent male	.492	.0101	.466	.569
Percent employment	.457	.0574	.213	.617
Income	$23,992	$5,691	$11,044	$55,157

Note.—Demographic information is not available on a monthly basis. Percent urban and percent uninsured are defined at the experiment-DMA level only, using data from 2000, because these variables are available only every 10 years with the census. All other demographic variables are defined at the experiment-DMA-year level.

sential. In this study, I do not compute consumer welfare, and the latent variables likely contain a combination of patient and physician utility, information, and persuasion. The purpose of the choice model is to guide the firm decision problem. Although these parameters might be related across levels by some kind of summing-up identity (as they would if each of the equations were only utility and consumers maximized utility), I do not restrict them to be related, because discovering the relative magnitudes of advertising effects at each level is a main question of this study.

I also note that this model incorporates unobserved heterogeneity through the inclusion of fixed effects, both for the market and for the comparison group time effect. That is, different effects might arise for each market and time period, and the average of those effects will be the reported coefficient on advertising. The model also incorporates observed heterogeneity in advertising effects using demographic information from the census. In particular, I use the percent black, percent Hispanic, percent Asian, income, percent uninsured, percent over age 45, and the employment-to-population ratio.[12] Including heterogeneity over and above these fixed effects and demographic interactions resulted in no significant findings, perhaps because these fixed effects explain so much variation. Descriptive statistics for the demographic variables are available in table 3. For ease of interpretation, when included in the demand model, these demographics are log normalized with mean zero and a standard deviation of one.

[12] These demographic interactions are the same as those included in Stremersch et al. (2013) plus percent uninsured and employment-to-population ratio.

For general intuition of the model, consider what happens if a single product, Zoloft, raises advertising in a market while everything else remains constant. That advertisement may have three effects. First, it may raise the probability that a consumer purchases any antidepressant. That effect is expressed through the top-level equation, increasing a_{lmt}, which increases A_{lmt}, which in turn increases $\Gamma_1(A_{lmt})$. Next, the information in the advertisement may push the consumer toward the subcategory of antidepressants that Zoloft is in over another, because the commercials often contain information about mechanisms and side effects, which are highly correlated within subcategory. The Zoloft advertisement increases a_{nmt}, which increases A_{nmt}, which in turn increases $\Gamma_2(A_{nmt})$. The marginal revenue will depend on the shape of the curve and the amount of advertising by other products in the same subcategory. Finally, the advertisement may have a pure business-stealing effect. By increasing a_{jmt}, A_{jmt} and $\Gamma_3(A_{jmt})$ increase to take share away from other products within the subcategory.

Also note that the model does not explicitly examine the various possibilities for forward-looking consumers, including consumer learning, as in Crawford and Shum (2005), Ching (2010), or Dickstein (2014). However, the difference in persistence parameters from the category level to the product level allows consumers to purchase one brand, decide it does not work satisfactorily, and move to another brand. In particular, if the persistence parameter for category-level advertising is higher than that of product-level advertising, the consumer might still be consuming in the category but no longer the same product. Further, the product-time fixed effects allow me to take into account differences in market conditions. That is, in the year prior to Prozac going off patent, consumers who know they will be taking the drug for a while may wish to be prescribed Prozac rather than Zoloft, because they know a cheaper, chemically identical generic will be available in the following period that they could switch to easily. This effect is absorbed into the product-time effects if it exists.

B. *Derivatives and Elasticities*

Given product shares in equation (4) and the logit structure, we can get the derivative of s_j, which is in subcategory n with respect to new advertising, a_k, of product k, which is in subcategory n', by using the chain rule and the typical logit derivatives:

$$\frac{\partial s_j}{\partial a_k} = s_{j|n}\left(s_{n|l}\frac{\partial s_l}{\partial a_k} + s_I\frac{\partial s_{n|l}}{\partial a_k}\right) + s_{n|l}s_l\frac{\partial s_{j|n}}{\partial a_k}. \tag{11}$$

Solving this out using our specification on shares, we get derivatives,

$$\frac{\partial s_j}{\partial a_k} = \begin{cases} s_j\left[\frac{\partial \Gamma_1}{\partial a_k}(1 - s_l) + \frac{\partial \Gamma_2}{\partial a_k}(1 - s_{n|l}) + \frac{\partial \Gamma_3}{\partial a_k}(1 - s_{j|n})\right] & j = k \\ s_j\left[\frac{\partial \Gamma_1}{\partial a_k}(1 - s_l) + \frac{\partial \Gamma_2}{\partial a_k}(1 - s_{n|l}) - \frac{\partial \Gamma_3}{\partial a_k} s_{k|n}\right] & j \neq k \;\&\; n = n' \\ s_j\left[\frac{\partial \Gamma_1}{\partial a_k}(1 - s_l) - \frac{\partial \Gamma_2}{\partial a_k} s_{n'|l}\right] & j \neq k \;\&\; n \neq n', \end{cases} \tag{12}$$

and advertising elasticities equal to

$$\eta_{jk} = \begin{cases} a_k\left[\frac{\partial \Gamma_1}{\partial a_k}(1 - s_l) + \frac{\partial \Gamma_2}{\partial a_k}(1 - s_{n|l}) + \frac{\partial \Gamma_3}{\partial a_k}(1 - s_{j|n})\right] & j = k \\ a_k\left[\frac{\partial \Gamma_1}{\partial a_k}(1 - s_l) + \frac{\partial \Gamma_2}{\partial a_k}(1 - s_{n|l}) - \frac{\partial \Gamma_3}{\partial a_k} s_{k|n}\right] & j \neq k \;\&\; n = n' \\ a_k\left[\frac{\partial \Gamma_1}{\partial a_k}(1 - s_l) - \frac{\partial \Gamma_2}{\partial a_k} s_{n'|l}\right] & j \neq k \;\&\; n \neq n'. \end{cases} \tag{13}$$

From these equations, we can see that firm benefits from own advertising may flow through expansion of the category, as is denoted by the term $s_j(\partial\Gamma_1/\partial a_j)(1 - s_l)$, through expansion of the subcategory in $s_j(\partial\Gamma_2/\partial a_j)(1 - s_{n|l})$, and through business stealing within the nest in $s_j(\partial\Gamma_3/\partial a_j)(1 - s_{j|n})$. Firm benefits from rivals' advertising in the same subcategory may flow through expansion of the category in $s_j(\partial\Gamma_1/\partial a_k)(1 - s_l)$ or through expansion of the subcategory in $s_j(\partial\Gamma_2/\partial a_k)(1 - s_{n|l})$, whereas this same advertising may hurt through business stealing within the subcategory in $-s_j(\partial\Gamma_3/\partial a_k)s_{k|n}$. Advertising from rivals in other nests may benefit the firm only through the expansion of the inside option but may hurt through expansion of the other subcategory at the expense of the firm's subcategory. Note that this structure fully allows for advertising that is a pure category expansion (i.e., if $\partial\Gamma_2/\partial a_j = \partial\Gamma_3/\partial a_j = 0$ for all j), for advertising that is pure business stealing (i.e., if $\partial\Gamma_2/\partial a_j = \partial\Gamma_1/\partial a_j = 0$ for all j), or anything in between, including cross-subcategory substitution. Rival advertising outside of the subcategory might also help more than inside of the subcategory if $\partial\Gamma_2/\partial a_j$ is sufficiently small and $\partial\Gamma_3/\partial a_j$ is sufficiently large or vice versa. What is restricted is that a firm's own advertisements may not help another firm more than it helps itself in elasticity terms. In the most extreme scenario, it is pure category expansion and helps all firms equally. Whether advertising provides positive or negative spillovers depends on the relative strength of the market expansion and the business-stealing channels and is a result of estimation rather than an assumption of the model.

Note that through the category expansion channel, rivals' advertising moves a firm's marginal revenue with respect to advertising downward. However, own advertising must move a firm's residual marginal revenue curve even further downward, as decreasing returns occur at the conditional share level as well. Assuming that the effect of advertising is positive at all levels, the primary effect of own advertising is stronger than that of rivals' advertising, and decreasing returns to own advertising are more severe than decreasing returns to rival advertising. These implications are consistent with findings in the reduced-form section. In particular, the fact that the coefficient on the own advertising squared term is more negative than the coefficient on the rival advertising squared term leads to the first implication, and the fact that the coefficient on the cross partial is negative leads to the second.

V. Empirical Specification and Estimation of the Model

A. *Demand Specification*

I define the advertising stock at each level s, where $s \in \{l, n, j\}$ is either the category level, subcategory level, or product level, to be a lag of a nonlinear function of current advertising, similarly to Dubé et al. (2005):

$$A_{smt} = \sum_{\tau=0}^{t} \lambda_s^{t-\tau} \log(1 + a_{sm\tau}). \tag{14}$$

Specifying advertising stock as a concave function of each period's advertising allows the firm's problem to have a well-behaved optimum. I explored other functional forms, and none changed the results in any significant way.

The advertising stock enters into the utility specification linearly at each level:

$$\Gamma_s(A_{smt}) = \gamma_s A_{smt}. \tag{15}$$

I account for all product characteristics other than advertising with a rich set of fixed effects, as the only pieces of data that vary at the choice level, DMA, and time level are shares and advertising.

Substituting equations (14) and (15) into equations (3)–(5) for the market-level l obtains

$$u_{ilmt} = \gamma_l \left[\sum_{\tau=0}^{t} \lambda_l^{t-\tau} \log(1 + a_{lm\tau}) \right] + \alpha_{lt} + \alpha_{lm} + \xi_{lmt} + \varepsilon_{ilmt}. \tag{16}$$

The conditional utilities for the subcategory and product levels are defined analogously. From here, note that current-period advertising enters the utility function in a concave manner, so the firm maximization problem is well behaved.

B. *Transforming to a Linear Problem*

Following Berry (1994), at each level of the problem, I specify an "outside good," take the log of the market share, and subtract from it the log of the outside option share, thus resulting in a linear form.

At the category level, the outside good is naturally defined as the population not filling a prescription for an antidepressant in month t in market m:

$$\log(s_{lmt}) - \log(s_{0mt}) = \gamma_1\left[\sum_{\tau=0}^{t}\lambda_l^{t-\tau}\log(1 + a_{lm\tau})\right] + \alpha_{lt} + \alpha_{lm} + \xi_{lmt}. \quad (17)$$

At the subcategory level, the outside good will be defined as the subcategory of older-style TCA antidepressants. The share of a subcategory conditional on being in the inside option follows

$$\log(s_{nmt|l}) - \log(s_{0mt|l}) = \gamma_2\left[\sum_{\tau=0}^{t}\lambda_n^{t-\tau}\log(1 + a_{nm\tau})\right] + \alpha_{nt} + \alpha_{nm} + \xi_{nmt}. \quad (18)$$

At the product level, the outside option in each nest will be the set of all products that never advertise on television. The product share equation conditional on already having chosen subcategory n is

$$\log(s_{jmt|n}) - \log(s_{0mt|n}) = \gamma_3\left[\sum_{\tau=0}^{t}\lambda_p^{t-\tau}\log(1 + a_{jm\tau})\right] + \alpha_{jt} + \alpha_{jm} + \xi_{jmt}. \quad (19)$$

Now, using these equations to solve for inside option shares in time $t-1$ and substituting that expression back into the expression for time t shares yields

$$\begin{aligned}\log(s_{lmt}) - \log(s_{0mt}) = {} & \lambda_l[\log(s_{lm,t-1}) - \log(s_{0m,t-1})] + \gamma_1\log(1 + a_{lmt}) \\ & + \theta_{lt} + \theta_{lm} + \nu_{lmt},\end{aligned} \quad (20)$$

where

$$\theta_{lt} = \alpha_{lt} - \lambda_l\alpha_{l,t-1} \quad (21)$$

is an inside-option-time-specific taste or quality parameter, and

$$\theta_{lm} = \alpha_{lm} - \lambda_l \alpha_{lm} \tag{22}$$

is the category-market-specific taste parameter. Finally,

$$\nu_{lmt} = \xi_{lmt} - \lambda_l \xi_{lmt} \tag{23}$$

is a market-time-specific demand shock. Equation (20) is precisely a lagged dependent variable with fixed-effects specification as described above, making possible the use of the border identification strategy.

Similarly, subcategory- and product-level share equations may be specified as

$$\begin{aligned} \log(s_{nmt|l}) - \log(s_{0mt|l}) = {} & \lambda_n[\log(s_{nm,t-1|l}) - \log(s_{0m,t-1|l})] + \gamma_2 \log(1 + a_{nmt}) \\ & + \theta_{nt} + \theta_{nm} + \nu_{nmt}, \end{aligned} \tag{24}$$

$$\begin{aligned} \log(s_{jmt|n}) - \log(s_{0mt|n}) = {} & \lambda_p[\log(s_{jm,t-1|n}) - \log(s_{jm,t-1|n})] + \gamma_3 \log(1 + a_{jmt}) \\ & + \theta_{jt} + \theta_{jm} + \nu_{jmt}. \end{aligned} \tag{25}$$

C. *Identification and Estimation Strategy*

Because the share equations have been transformed to a linear form, estimation may be done by ordinary least squares. A notable problem in estimating this equation is that advertising is a firm choice variable determined in equilibrium and is thus endogenous. As such, I will take advantage of the discrete nature of DMAs to make use of spatial variation as described in Section III.

In particular, I specify the estimation equation as

$$\begin{aligned} \log(s_{lmt}) - \log(s_{0mt}) = {} & \lambda_l[\log(s_{lm,t-1}) - \log(s_{0m,t-1})] + \gamma_1 \log(1 + a_{lmt}) \\ & + \theta_{lbq} + \theta_{lbm} + \nu_{lbmt}, \end{aligned} \tag{26}$$

where θ_{lbq} is a border-time fixed effect and θ_{lbm} is a border-DMA fixed effect. Partialing these fixed effects out makes the identifying variation at the market level the total advertising in market m that is over and above the average on its side d of the border b and over and above the average local total advertising in quarter q in all counties on either side of border b. The fixed effects will also control for the product quality terms θ_t and θ_m.

I identify the effects at the other two levels similarly. In the subcategory level, I include fixed effects α_{nbq} and α_{nbm}, and at the product level, I include fixed effects α_{jbq} and α_{jbm}. Identifying variation will come at the subcategory level from total subcategory advertising that is above and be-

TABLE 4
DESCRIPTIVE STATISTICS: BORDER SAMPLE, 1997–2003

	Mean	Quarter 10	Median	Quarter 90
Number of border experiments	153			
Number of DMAs	97			
Log $DTC_{product}$	.190	0	0	1.033
Log DTC_{nest}	.447	0	0	1.682
Log DTC_{market}	.817	0	.921	1.987

NOTE.—Log DTC equals the log of one plus DTC expenditures per 100 capita. All are defined at the experiment-DMA-month level.

yond the historical advertising in its market and above the border average in the current time period. At the product level, identifying variation will be advertising for product *j* that is above and beyond advertising for product *j* on the border in quarter *q* and above and beyond the average over all time in market *m*. No between-product variation in advertising will be used to identify the advertising parameter.

Table 4 has variable definitions and summary statistics for those variables that will enter the estimation.

D. *Demand Results*

1. Effects at Each Level

Results are presented in table 5. The effect of advertising stock on demand at each stage of the decision is positive. The strongest effects are at the category level, deciding between the inside and outside option, and at the product business-stealing level. Effects at the subcategory level are not significant, but note that there is advertising in only two subcategories, with most of the advertising happening in the SSRI subcategory. The small and insignificant effect at the subcategory level is not surprising, because patients having good information about what separates the subcategories seems unlikely. The demographic interactions largely show insignificant results. At the category level, areas with a higher population over age 45 have a slightly higher advertising effect. At the product level, the business-stealing effect is stronger in areas with a higher percentage of females. All other specifications in the paper will include these demographic interactions but will suppress them, because their inclusion or exclusion does not affect the estimated main effects.

Table 6 presents short-run demand elasticities of current advertising, showing that the category expansive properties of advertising dominate the business-stealing effects and all cross-advertising elasticities are positive. This finding is consistent with the identified positive spillovers in the reduced form.

TABLE 5
Results of Base Model

Variable	Category Level	Subcategory Level	Product Level
Ad stock	.0496***	.00694	.0254**
	(.00793)	(.00834)	(.00912)
× percent black	−.00255	.00544	−.00392
	(.00892)	(.0118)	(.0103)
× percent Hispanic	−.0172	−.0159	.0123
	(.0151)	(.0144)	(.0127)
× percent Asian	.00676	−.0101	−.00229
	(.0114)	(.0168)	(.0130)
× percent urban	−.00204	.00353	−.00269
	(.00788)	(.0119)	(.0162)
× percent uninsured	.0137	−.00130	.000723
	(.0111)	(.0136)	(.0102)
× percent over 45	.0149*	.0195	−.00651
	(.00712)	(.0111)	(.0103)
× percent male	−.00334	.00930	−.0197***
	(.00789)	(.00737)	(.00689)
× employment	.00684	−.0146	−.00276
	(.00850)	(.0117)	(.0110)
× income	−.00662	.0185	−.00957
	(.00987)	(.0152)	(.0116)
Persistence, λ	.680***	.279***	.324***
	(.0306)	(.0120)	(.0139)
Observations	22,592	92,238	139,305
R^2	.952	.936	.961

Note.—Level-DMA clustered standard errors are in parentheses. Demographic variables are log normalized with mean zero and standard deviation of one for ease of interpretation. Level-border-DMA and level-border-quarter fixed effects are included to execute the border strategy as described in the text.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

2. Persistence

Persistence is highest at the category level. The persistence parameter of 0.68 implies that 90 percent of the effect dissipates within 6 months. Meanwhile, the 0.32 persistence parameter at the product level implies

TABLE 6
Short-Run Advertising Elasticities

Product	Paxil	Paxil CR	Prozac	Prozac Weekly	Wellbutrin SR	Wellbutrin XL	Zoloft	Outside Option
Paxil	.037	.019	.021	.019	.020	. . .	.021	−.023
Paxil CR	.016	.029	.016	.016	.012	.010	.016	−.015
Prozac	.0092	. . .	.020	.0080	.0097	. . .	.0092	−.011
Prozac Weekly	.0088	. . .	.0088	.018	.0068	. . .	.0088	−.0080
Wellbutrin SR	.014	. . .	.014	.012	.021	. . .	.014	−.014
Wellbutrin XL	.017	.017	.017	.017	.019	.035	.017	−.018
Zoloft	.013	.013	.013	.013	.013	.010	.027	−.015

that 90 percent of the business-stealing effect of an advertisement dissipates within only 2 months. This result suggests that advertising in the long run is more of a category expansion than a business-stealing tool. This finding is consistent with the common wisdom that antidepressants are subject to a high degree of experimentation. If a patient tries one and finds the side effects unbearable, she might well switch to another one rather than quit antidepressants altogether. The finding is also consistent with a limited memory view of advertising. Because advertisements for pharmaceuticals on television usually contain a lot of information about the condition, the mechanisms of action, and the side effects and these characteristics are highly correlated within category, a consumer might well remember seeing an advertisement about depression without remembering which brand was advertised. This high persistence at the category level relative to the product level is another source for potential underinvestment in advertising relative to a cooperative.

3. Importance of Using the Border Strategy

Table 7 presents estimates that highlight the importance of using the border strategy to account for the endogeneity of firm choice. In column 1, the demand analysis occurs at the DMA level without using market or time fixed effects. Market fixed effects control for persistent differ-

TABLE 7
Main Results: Importance of the Border Strategy

Variable	No Border (1)	No Border (2)	Border (3)
Ad $\text{stock}_{\text{category}}$	.0118***	.0453***	.0496***
	(.00113)	(.00234)	(.00793)
$\lambda_{\text{category}}$	.968***	.715***	.680***
	(.00523)	(.0202)	(.0305)
Ad $\text{stock}_{\text{subcategory}}$	.0326***	.00243	.00694
	(.00221)	(.00324)	(.00834)
$\lambda_{\text{subcategory}}$	.973***	.699***	.279***
	(.00122)	(.0138)	(.0120)
Ad $\text{stock}_{\text{product}}$	.0652***	.0203***	.0254**
	(.00669)	(.00524)	(.00912)
λ_{product}	.961***	.594***	.324***
	(.00184)	(.0115)	(.0139)
Fixed effects		X	X

Note.—Level-DMA clustered standard errors are in parentheses. Demographic variables are log normalized with mean zero and standard deviation of one for ease of interpretation. Level-border-DMA and level-border-quarter fixed effects are included to execute the border strategy as described in the text.

* $p < .05$.
** $p < .01$.
*** $p < .001$.

ences in demand patterns, whereas the time fixed effects control for national market conditions, such as patent expirations, new clinical study results, and new product introductions. Column 2 includes these fixed effects in a DMA-level analysis but does not control for the endogeneity of firm choices using the border strategy. Column 3 uses the border strategy and fixed effects.

Notable from these results is that failure to use fixed effects understates the category-level effects and overstates the business-stealing effects of advertising. In addition, persistence parameters are drastically overstated. The naive analysis would suggest that advertising is extremely persistent and primarily business stealing. Why might this be? Firms are likely to target their advertisements to markets that generally have a higher preference for their products, leading to the overstatement of the business-stealing effect. Further, generic introductions likely skew advertising decisions. Branded products on the whole would prefer to advertise more in markets containing a lower taste for generics or low generic penetration for any other reason. Meanwhile, in markets with a higher generic penetration, total prescriptions could be higher because of the lower-priced generics. This combination of factors leads the researcher to understate the category expansive effects of advertising. Controlling for generic introductions and other market conditions using time fixed effects, as well as controlling for persistent market differences using DMA fixed effects, may mitigate both of these concerns.

Upon introducing the DMA and time fixed effects in column 2 of table 7, the point estimates on advertising fall more in line with those found using the border approach. However, the persistence parameters at both the subcategory and product levels are significantly overstated. This finding is consistent with a firm rule-of-thumb strategy, whereby firms set advertising as a multiple of previous-period demand. In that case, the lagged dependent variable will control for this rule of thumb, making it not only a measure of state dependence but also a measure of firm targeting. As such, the persistence parameters are overstated. Given that the overstatement of these persistence parameters is more severe at the subcategory and product levels, the researcher will not only overstate the long-run effectiveness of advertising but also understate the magnitude of the spillovers over time.

VI. Supply and Counterfactual

A. *Supply Implications of Positive Spillovers*

The demand results above imply that the incentive to invest in advertising is dampened by positive spillovers for two reasons. First, advertising provides benefits to rival firms that the advertising firm does not internalize. Second, rival advertising lessens the incentive to advertise by lowering the marginal category expansive effect of advertising.

1. Internalization Incentives

To further illustrate the effects of advertising over time on rivals, consider an impulse response graph in figure 8. The purpose of this graph is to follow the effect of a marginal dollar per 100 capita spent by Zoloft in January of 2002 on both Zoloft and total market prescriptions for the subsequent year. The top downward-sloping curve is the marginal effect of Zoloft advertising on total market prescriptions, and the bottom downward-sloping curve is the marginal effect of Zoloft advertising on Zoloft prescriptions. The upward-sloping dashed line is the ratio of the total market effect to the Zoloft effect. A marginal dollar per 100 capita of Zoloft advertising leads to a contemporaneous increase of about 70,000 antidepressant prescriptions, only about 20,000 of which are captured by Zoloft. Further, as we follow that marginal dollar through time, the effect on the total market is more persistent, and the marginal effect of Zoloft advertising goes more and more to other products. A large contemporaneous positive spillover occurs that intensifies through time. Zoloft has no incentive to internalize the benefits it bestows on other

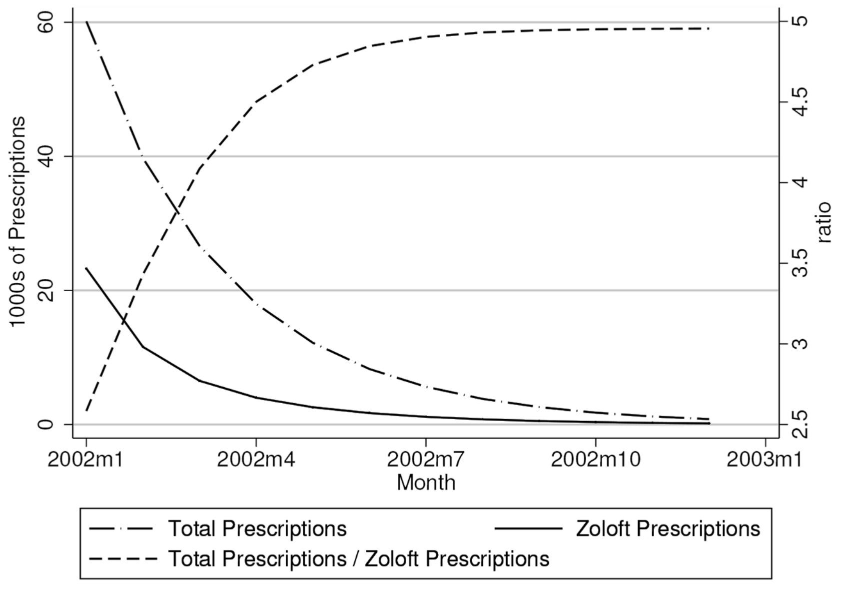

Fig. 8.—Impulse response effect of Zoloft advertisement on own and total prescriptions

firms and thus will underinvest in advertising relative to a cooperative controlling advertising in the whole market.

2. Free-Riding Incentives

To illustrate the free-riding incentive, I consider three marginal revenue curves and a horizontal marginal cost curve in figure 9 given the demand parameters estimated in the previous section, but for a single point in time and for a single product. In the figure, I consider the perspective of Zoloft in January 2002 in the Boston DMA.

The top curve is the marginal revenue for Zoloft if all competitors set advertising equal to zero. Notably far below that curve, the middle curve is the marginal revenue curve of Zoloft if competitors combine to advertise \$3 per 100 capita, which is about the average competitor advertising Zoloft sees in the Boston DMA during the time it advertises. Finally, the lowest curve depicts the marginal revenue with respect to advertising of Zoloft when its competitors advertise \$10 per 100 capita, about the maximum it ever faces from competitors in the Boston market. Note from the curves that the marginal revenue curve of Zoloft takes a significant hit as its competitors advertise more. In fact, when competitors advertise up to \$10 per 100 capita, advertising at all is almost not worthwhile for Zoloft. Zoloft has a clear incentive to free ride as competitors advertise more and more.

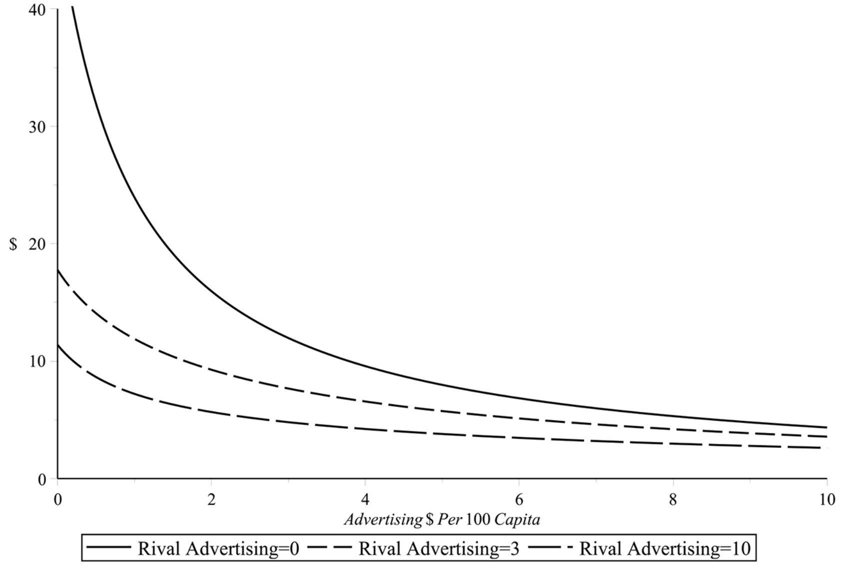

FIG. 9.—Marginal revenue curves under various scenarios

3. Is Firm Behavior Consistent with Free Riding?

Whereas the last two sections show that the demand estimates imply an incentive for firms to underinvest through internalization failures and free riding, they do not provide evidence of actual firm free riding. To evaluate whether firm behavior is consistent with free riding, we can test whether firms engage in less DTC advertising in markets where rivals engage in large amounts of DTC advertising, controlling for market and time effects. That is, when a given market gets more rival DTC than is typical in that market, is the firm less likely to engage in DTC in that market at that time? To address these questions, I estimate four regressions of the form

$$f(a_{jmt}) = g\left(\sum_{-j} a_{-jmt}\right) + \alpha_{jm} + \alpha_{jt} + \varepsilon_{jmt}, \tag{27}$$

where f is an increasing function of own-product advertising and g is a function of rival product advertising. Table 8 addresses this question in four different ways. In columns 1 and 2, the amount of advertising is evaluated. Conditional on the market and time effects from the model, do firms advertise less in markets where rivals advertise more? In column 1, $f(a_j) = a_j$ and $g(\Sigma a_{-j}) = \Sigma a_{-j}$. Given this specification, firms advertise less when their rivals advertise more. In particular, an additional \$1 per 100 capita by a rival is associated with a decrease in own advertising of about \$0.03. This effect is small but significant. In column 2, the same question is evaluated in logs, where $f(a_j) = \log(a_j)$ and $g(\Sigma a_{-j}) = \log(\Sigma a_{-j})$. When rivals increase advertising by 10 percent, firms decrease advertising by 0.3 percent. Columns 3 and 4 address the extensive margin. Are firms less likely to advertise at all in a market if rivals advertise more? In column 3, $f(a_j) = \mathbf{1}(a_j > 0)$ and $g(\Sigma a_j) = \mathbf{1}(\Sigma a_{-j} > 0)$. The pres-

TABLE 8
Evidence of Free Riding by Firms

Variable	DTC	Log(DTC)	$\mathbf{1}$(DTC > 0)	$\mathbf{1}$(DTC > 0)
Rival DTC	−.0296***			−.00334*
	(.00240)			(.00160)
Log(rival DTC)		−.0290***		
		(.00691)		
$\mathbf{1}$(rival DTC > 0)			−.0685***	
			(.0162)	
Observations	39,414	39,414	39,414	39,414

Note.—Product-DMA clustered standard errors are in parentheses. These regressions are run only for those products that ever advertise on television. Product-DMA and product-quarter fixed effects are included as in the DMA-level demand analysis.

* $p < .1$.
** $p < .05$.
*** $p < .01$.

ence of rival advertising in the market decreases a firm's likelihood of advertising in that market by about 0.068. Finally, in column 4, $f(a_j) = \mathbf{1}(a_j > 0)$ and $g(\Sigma a_{-j}) = \Sigma a_{-j}$. As rivals increase DTC by $1 per 100 capita, a firm's likelihood of advertising in that market decreases by 0.003. In all specifications, rival advertising is negatively and statistically significantly associated with both the amount of firm advertising and the decision whether to advertise, which is consistent with firms having some understanding of the free-riding incentives generated by spillovers in this market.[13]

B. Supply Simulation

1. Purpose

Although the previous section shows that firm behavior is consistent with free riding, it does not speak to the size of the free-riding or internalization incentives implied by the demand model. In this section, I will use a highly stylized model to illustrate this point. To be clear, the purpose of this model is not to predict the data, but to show the theoretical magnitude of the incentive effects of spillovers holding all other factors fixed. I will do this exercise by plugging the parameters from the demand model into a stylized equilibrium model whereby the firm may adjust DTC but must hold pricing, detailing, and other factors fixed. I will then compare this oligopoly benchmark with a counterfactual scenario whereby firms cooperate on advertising. This comparison will provide an upper bound on the magnitude of the dynamic underinvestment in advertising due to the positive spillovers, because the oligopoly benchmark assumes firms are optimally free riding and failing to internalize their benefits on rivals.

Predicting the observed data using this type of model is complicated by the fact that firms at the time of the sample would need to have a level of sophistication high enough to know the exact effects of advertising. In particular, there is evidence suggesting that firms in fact did not have this level of sophistication at the time of the sample. In 2002, the Association of Medical Publications commissioned the Analysis of Return on Investment for Pharmaceutical Promotion (ARPP) study (Wittink 2002). This study had a steering committee including representatives from Wyeth,

[13] Although these results are suggestive of and consistent with free-riding behavior, they are not definitive. These results are also consistent with other potential stories. For example, firms could believe that customers have limited attention with respect to watching television advertising for antidepressants. If rival firms commit to advertising this month, a firm might prefer to wait to advertise until next month, even if the rival advertisements have a negative impact on own demand. Other scenarios might also lead own and rival advertising to be strategic substitutes as well.

GlaxoSmithKline, Bayer, and Novartis. The study was meant to suggest to firms how they might use data analysis to infer the relative usefulness of various elements of marketing mix. Among other things, the study discusses the use of pooled regression analysis. That the pharmaceutical firms were involved in such a study suggests they might not have a complete idea as to how to optimally set advertising. Even if the model were able to predict the observed data, firms might well respond very differently to out-of-sample counterfactual scenarios. For example, in the data, firms do not respond to DTC by increasing detailing, but that lack of response is no guarantee that they would not do so in response to a large joint advertising campaign. As such, this simulation exercise is not meant to be predictive, but rather to illustrate the economic magnitude of the incentives generated by the demand system if firms could optimally free ride holding other factors fixed.

2. Cautions and Caveats in Interpreting the Supply Side

It is very important to note that firms might not be playing a dynamically optimal advertising strategy as is assumed in the stylized model to follow. Indeed, I will show that observed firm strategies do not match what the optimal dynamic oligopoly would suggest. Firms could be myopic or boundedly rational as a result of managerial short-sightedness, costs of measuring advertising effectiveness, or a lack of quantitative training for marketing managers. As such, the quantitative results from this section should be interpreted with caution.

However, given that the estimated demand system shows positive spillovers and strategic substitutability in advertising, firms will underinvest in advertising relative to an industrywide cooperative in almost any model in which firms optimize either their static or dynamic profits with respect to own advertising. Even if firms do not act strategically with respect to rival advertising, they will underinvest in advertising because of an internalization failure: they will only maximize their own profits with respect to advertising rather than considering the profits their advertising might bring to rivals. If firms act strategically with respect to rival advertising, they will further underinvest as a result of free riding off of their rivals' efforts.

With those caveats in mind, I choose to measure the size of the potential underinvestment assuming that firms dynamically optimize with respect to advertising and act strategically with respect to rival advertising. While firms in this sample from 1999–2003 do not appear to behave in a way consistent with dynamic optimality, we might expect them to going forward for two reasons. First, this study provides evidence of both a long-run impact of advertising on demand and positive spillovers. That evidence may not have been available in the late 1990s and early 2000s.

Given the new information, firms might alter their behavior to improve their performance. Second, marketing analytics has become far more sophisticated in the past decade. Companies have significantly expanded their marketing analytics groups, and a new field of data science has produced many highly quantitative managers. Given the new evidence along with increasingly rigorous quantitative training, firms may become more sophisticated over time. Hence, even though a dynamic game model (Markov perfect equilibrium) may not be the right model for the sample period, it seems plausible that it could be a reasonable model in the foreseeable future.

3. Model Setup

The model will be based on a Markov perfect equilibrium (MPE) concept that allows firms only to optimize over DTC advertising. Firms will solve a dynamic programming problem to incorporate both present and future payoffs associated with advertising. I use this framework for two main reasons. First, solving the dynamic programming problem allows the evaluation of the total future discounted "marginal revenues" associated with advertising today. Second, not allowing firms to adjust prices or detailing highlights the theoretical size of the effect of advertising spillovers on incentives to advertise. The problem will assume that firms have a monthly discount rate of 0.95 and a marginal cost of advertising $1.00 of $1.15, including both the pecuniary cost of advertising and a 15 percent agency fee. In this way, all inputs into the model are observable, and all that remains is to simulate the equilibrium. Details of the formulation of the framework are provided in appendix F. Having solved this equilibrium as a benchmark, I then compute "counterfactual" scenarios whereby the advertising firms work together in an advertising cooperative to internalize the positive spillovers and eliminate free riding between them. The difference between the benchmark and the counterfactual outcomes will illustrate the size of the underinvestment incentives.

One complication of solving an MPE in this setting is that the market conditions change frequently. That is, products enter throughout the sample, both from new innovation and from patent expiration, which leads to generic entry. Similarly, with patent expiration, products effectively leave the market (i.e., their market shares become very low and they cease to participate in the advertising game). In the demand estimation, I deal with these issues using product-time fixed effects. In principle, we might view these effects as additional state variables. Because entry and exit are not the main focus of this study, my strategy will be to focus on a period within my sample during which the market is stable. From December 1999 until December 2000, no product entries or exits occur. The next major product entry occurs in July 2001, when Prozac goes off patent and generic Prozac enters. Indeed, over the year from December 1999 until December 2000, the estimated product-time fixed effects for the advertised products remain relatively constant. Further, because the effects of advertising almost entirely dissipate over the course of 7 months, the infinite horizon dynamic programming problem assuming stable market conditions from December 1999 through December 2000 is a reasonable approximation to one that includes the important market changes in July 2001.

For the period described, only two products advertise on television: Paxil and Prozac. Paxil is produced by GlaxoSmithKline, which also produces Wellbutrin and Wellbutrin SR during this period. Prozac is produced by Eli Lilly, which produces only Prozac in this window. As such, Paxil will advertise in order to maximize discounted future profits over Paxil, Wellbutrin, and Wellbutrin SR, while Prozac will advertise to maximize discounted future profits for Prozac. During this time period, Prozac has a larger market share than Paxil and, in fact, has the largest market share of any antidepressant. As such, it will have the largest incentive to advertise.

Prices and marginal costs of production are given in the data. For a reference point, the average price of a prescription of Paxil over the course of the period is \$65, while the average production cost is \$3.68. The average price for a prescription of Prozac is \$72 and the production cost is \$3.53. Details of the computation of the equilibrium are in appendix F. The profit-maximization problem is concave, so an equilibrium is found relatively easily.

For the sake of exposition, results are shared for the Atlanta DMA. In particular, prior to the period in question, only Paxil advertised. That fact is included as an initial condition. Figure 10 shows the evolution of advertising for Paxil and Prozac over the course of the period in the benchmark scenario. Paxil begins the period advertising about \$2.00 per 100 capita and increases its advertising to about \$3.00 per 100 capita by the third month, holding constant for the remainder of the period. Prozac advertises just under \$8.00 per 100 capita in the first period and increases advertising up to nearly \$9.00 per 100 capita by the seventh period and holds constant from there. Given these levels of advertising, the category share rises from an initial condition of 4.9 percent of the population to about 6.7 percent of the population over the course of the year. Profit evolution for each firm in the benchmark scenario is provided in figure 11. Eli Lilly's profits increase from \$3.2 million per month in Atlanta in the pre period to about \$4.1 million per month by the end of the period, while GlaxoSmithKline's profits (the combined profits of Paxil, Wellbutrin, and Wellbutrin SR) go from about \$2.7 million per month to about \$3.6 million per month. Combined, by the end of the sample,

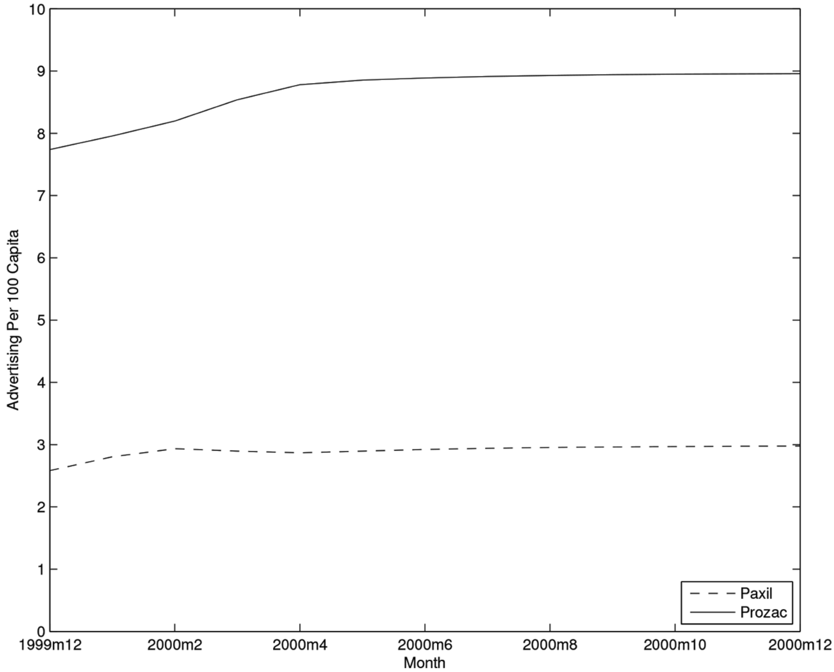

FIG. 10.—MPE simulation advertising—Atlanta DMA

GlaxoSmithKline and Eli Lilly are advertising about $11 per 100 capita and earning $7.7 million per month by the end of the period.

Comparing this benchmark MPE outcome with what is observed in the data (shown in fig. 12), Prozac advertises far less than would be predicted from the demand model, suggesting they are underinvesting more than the demand parameters would suggest they should. This finding could be a result of organizational costs, myopia, budget constraints, or simply a matter of learning the game, because DTC was relatively new to the pharmaceutical industry during this period. Paxil sometimes advertises a little more and sometimes a little less than the demand parameters would suggest they should. As mentioned, the purpose of this model is not to predict the observed levels of advertising but to illustrate the theoretical magnitude of incentive effects. Given the existence of the ARPP study, it might not be surprising that these numbers do not match the data.

4. Counterfactual Scenarios

I use two alternative scenarios to size the incentive effects of positive spillovers. First, I assume that Eli Lilly and GlaxoSmithKline cooperate on ad-

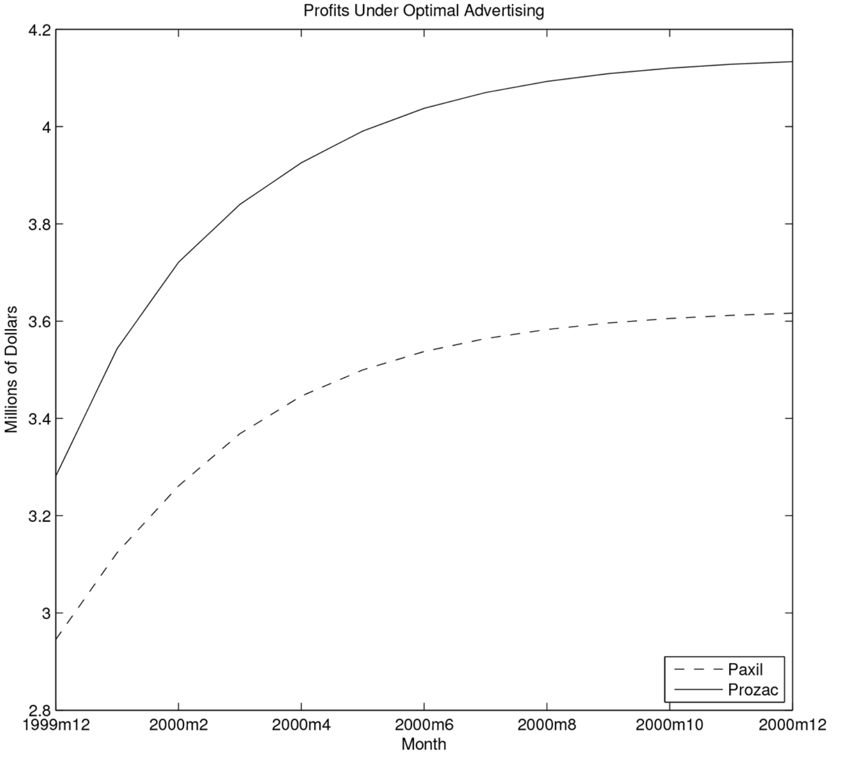

FIG. 11.—MPE simulation profits—Atlanta DMA

vertising that maximizes the combined profits of their products, which takes away the need for strategic response. Next, I will consider a scenario in which the entire market (rather than just the two firms observed to advertise) is allowed to set advertising in a single optimization problem to think about the potential of a category-wide advertising cooperative. The ability to contract on coordination would allow firms to overcome the free-riding problem and provide advertising, even without a brand-level component. As mentioned before, because the benchmark assumes firms are optimally free riding, we should think of this comparison as an upper bound on the magnitude of the dynamic underinvestment effect of the positive spillovers.

For the purposes of the counterfactuals, I assume that the advertising firms in the antidepressant market cooperate to make a common nonbranded category advertisement for antidepressants, facilitated by a patient advocacy group. Such groups are focused on educating patients on specific diseases and treatments. The specific mission of these types of organizations is to educate patients on how, when, and why to seek treat-

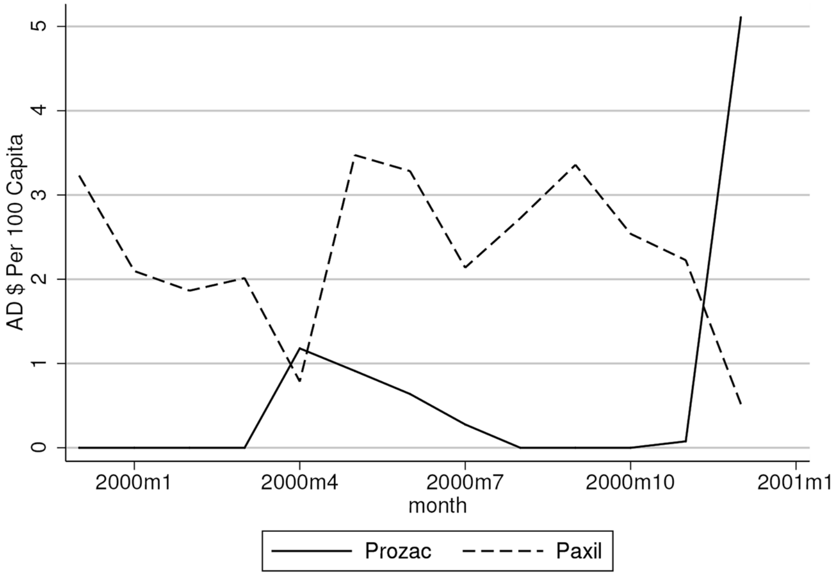

FIG. 12.—Realized advertising—Atlanta DMA

ment for various health care needs. Some examples of these types of organizations are the American Cancer Society, the Alzheimer's Association, and the Depression and Bipolar Support Alliance. While they do not tend to advertise on television currently, they might be an ideal facilitator for category-level advertising of antidepressants. The effect of those advertisements is assumed to be equal to the category-level effect of the branded advertisements estimated above. Because we observe none of these types of advertisements in the data, this effect is assumed. We could alternatively assume the cooperative agrees that only one product uses branded advertisements and then transfers the business-stealing rents via a prearranged contractual arrangement, and the analysis would follow in the same manner.

The cooperative solves the firm's problem in each month and in each market. In the Eli Lilly/GSK scenario, the cooperative includes Paxil, Prozac, Wellbutrin, and Wellbutrin SR as part of the portfolio. In the full cooperative scenario, all products in the category are included as part of the portfolio in the firm's problem. The marginal cooperative advertisement dollar has cost equal to \$1.15, as before. The cooperative solves the same optimization problem as in the benchmark, but because all products in the market are included in the optimization, strategic response is not necessary.

As mentioned previously, in these scenarios, firms are restricted from adjusting detailing or prices in response to these cooperative advertising campaigns for two main reasons. First, neither prices nor detailing is responsive to DTC in the data. This unresponsiveness is shown empirically in appendix A. The pricing result is also consistent with previous research that shows brand prices are tightly predicted by a product fixed effect and a time trend (Aitken et al. 2013). Second, not allowing prices or detailing to change highlights the incentive effects of advertising spillovers on advertising decisions. Note that whereas detailing is not associated with DTC in the data, a cooperative campaign is enough out of sample that firms may change their strategies once it arises. This stylized model cannot provide insight into that dimension of firm strategy. Also note that because, in the benchmark scenario, firms underinvest in advertising, the counterfactual scenarios will result in out-of-sample values of advertising. The functional form assumption in the model drives the exact magnitude of effects of out-of-sample advertising.[14]

C. *Advertising*

As the business-stealing incentive grows, benchmark total advertising is expected to increase relative to the counterfactual advertising. As the business-stealing incentive dwindles, the free-riding incentive associated with the positive spillovers should lead to lower benchmark advertising relative to the cooperative's ideal. As in the demand estimation, the business-stealing effects of advertising are swamped by the category expansive effects, and cooperation should lead to an increase in total category advertising.

For illustrative purposes, figure 13 shows the computed benchmark versus cooperative advertising choice in the Atlanta DMA in the cooperative scenarios described above, with the top panel being the Eli Lilly/GSK cooperative and the bottom panel being the full category advertising cooperative. The Eli Lilly/GSK cooperative advertises about 50 percent more in total than the sum of Prozac and Paxil in the competitive equilibrium. The full category cooperative advertises almost four times as much as the competitive equilibrium.

D. *Quantities and Profits*

Figure 14 illustrates the difference in profits between the benchmark and the cooperative situations. In the top panel, we see that in the Eli Lilly/GSK cooperative, the included products see roughly 10 percent higher profits than they do combined in the benchmark. In the bottom

[14] I tried alternative specifications with quadratic and square root functions of advertising, and they produced qualitatively and quantitatively similar results.

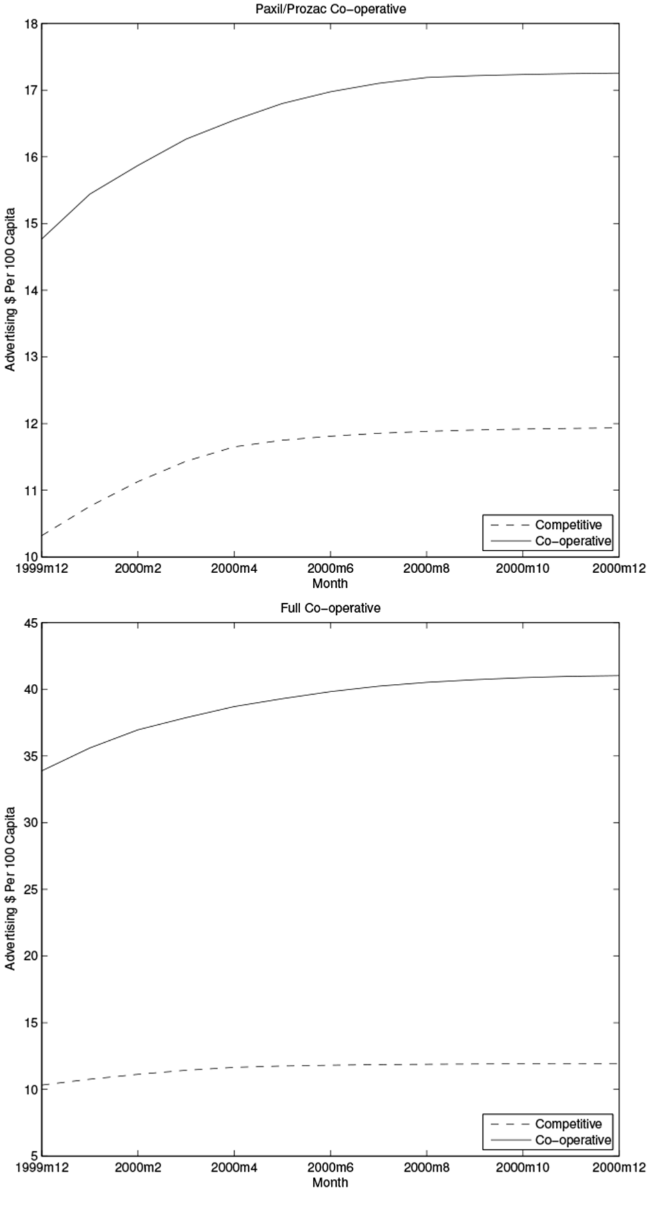

FIG. 13.—Paxil/Prozac and full cooperative versus competitive total advertising—Atlanta DMA.

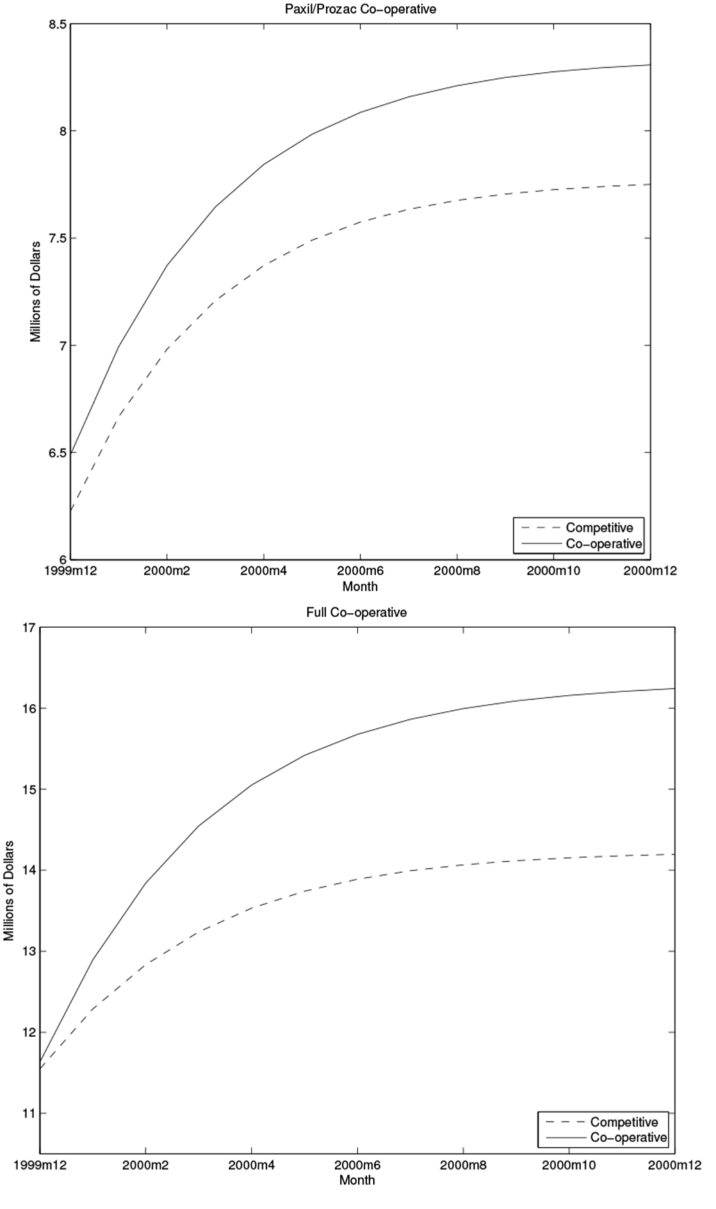

Fig. 14.—Paxil/Prozac cooperative versus competitive total profits—Atlanta DMA

panel, we see that the full category profits increase by about 14 percent by the end of the period. Figure 15 compares the category share of the population in the Atlanta DMA between the cooperatives and the benchmark. In the Eli Lilly/GSK cooperative, the category is about 6 percent

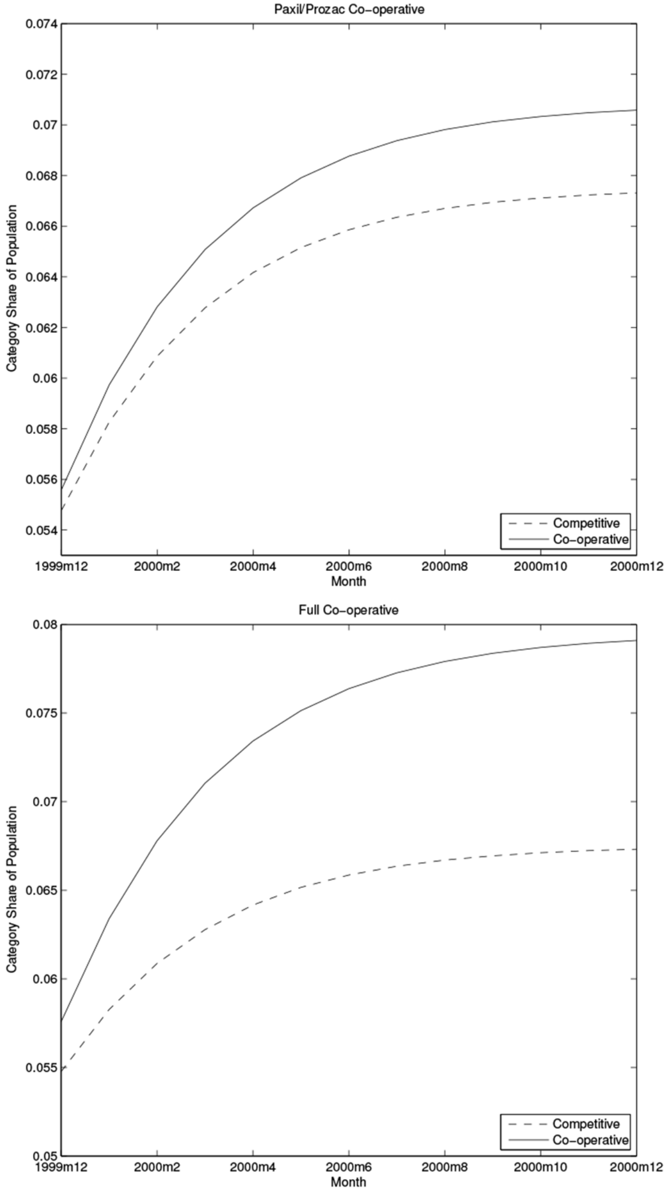

FIG. 15.—Paxil/Prozac cooperative versus competitive category share—Atlanta DMA

larger than it is under competition. In the full cooperative, the category is almost 18 percent larger by the end of the period than it is under competition.

E. Discussion

Although market expansion and increasing profits in the counterfactual would be viewed as welfare increasing in many consumer goods markets, we have several reasons to be cautious in drawing conclusions about social welfare in the context of antidepressants. First, many prescriptions are covered by insurance. Although many people are receiving prescriptions and incurring minimal, if any, cost, the insurance system pays out a significant price. The societal cost might not justify the new prescriptions. However, because many physicians see depression as an undertreated condition, the total welfare benefits could also be very high. Second, if I have missed important price or detailing complementarities, all increased profits might be competed away after the cooperative sets higher advertising, in which case, the welfare effect is also ambiguous. Finally, frictions within the firm might prevent optimal free riding, so these results should be interpreted as an upper bound on the dynamic incentive effects. The balance of societal costs and benefits, although interesting and important, is not identified in this study and is certainly worthy of further research.

VII. Conclusions

Using data from the antidepressant market, I find that television advertising has significant positive spillovers. I identify these effects using the discontinuity in advertising generated by the borders of television markets. The strategy proves important, because failing to consider endogenous firm choices of advertising leads to overstatements of the long-term effects, particularly of the business-stealing component of advertising. Consistent with the incentives generated by positive spillovers, I find that firms advertise less and are less likely to advertise at all in markets in which positive shocks to rival advertising occur. To provide a magnitude of the theoretical size of the incentives generated by these spillovers, I construct and simulate a model to systematically explore this fact and its implications on the supply decisions of firms. In particular, I find that the spillovers induce a free-riding and internalization problem whereby competitive advertising is significantly lower than the optimal strategy that a cooperative would set if it controlled the entire market. If the advertising firms worked together, they would advertise significantly more, increase the size of the category by 6 percent, and increase their own profits by 10 percent. Meanwhile, a full industry cooperative would set advertising four times as high as is observed in equilibrium and would increase industry shares by 18 percent and profits by 14 percent.

These findings also speak to some of the controversy surrounding the practice of advertising pharmaceuticals on television. Contrary to some of the criticism, this type of advertising drives consumers into the market and helps all products in the category, including the low-cost generics. It is the proverbial rising tide that lifts all ships. Although a brand effect is present, it is short-lived, whereas the category expansion effect is more persistent. Especially for conditions that are seen as undertreated, this type of advertising could be beneficial.

These findings are potentially relevant to firms, regulators, econometricians, and marketers. Firms might be able to realize gains from cooperation that regulators might allow. In the absence of cooperation, firms must properly take account of spillovers when deciding advertising policy. Regulators should take into account that content regulation might reduce or eliminate the firms' incentives to advertise. Finally, marketers and econometricians must consider the possibility of positive spillovers when building models of advertising impacts on supply and demand.

Limitations.—The current study has some limitations that future research could address. First, the magnitude of the free-riding and internalization underinvestment problem from the counterfactual should be viewed as an upper bound on the dynamic incentive effects of spillovers, because the benchmark competitive model assumes firms optimally free ride. If frictions within the firm prevented optimality, the increase in advertising due to cooperation might be smaller. Second, although the data show that firm behavior is consistent with free riding, they do not rule out other potential stories leading to own and rival advertising being strategic substitutes. More careful research to pin down how firms interact with each other in this market would be illuminating. Third, although both the national aggregate and physician sample panel data in appendix A show negligible correlation between DTC and detailing in aggregate and at the DMA level, we cannot be entirely certain the same would hold if we had detailing data for the full population of physicians. Including the full population of physician detailing data would provide a more convincing test of the parallel trends at the border assumption in detailing. Finally, all of the results rest on the assumption that all omitted variables, including national magazines and newspapers, follow parallel trends at the borders of DMAs. To the extent that these data are not available for the current study, the assumption is untestable.

References

Aitken, Murray L., Ernst R. Berndt, Barry Bosworth, et al. 2013. "The Regulation of Prescription Drug Competition and Market Responses: Patterns in Prices and Sales Following Loss of Exclusivity." Working Paper no. 19487, NBER, Cambridge, MA.

Anderson, Eric T., and Duncan Simester. 2013. "Advertising in a Competitive Market: The Role of Product Standards, Customer Learning, and Switching Costs." *J. Marketing Res.* 50 (4): 489–504.

Anderson, Ian M. 2000. "Selective Serotonin Reuptake Inhibitors versus Tricyclic Antidepressants: A Meta-Analysis of Efficacy and Tolerability." *J. Affective Disorders* 58 (1): 19–36.

Avery, Rosemary J., Matthew D. Eisenberg, and Kosali I. Simon. 2012. "The Impact of Direct-to-Consumer Television and Magazine Advertising on Antidepressant Use." *J. Health Econ.* 31 (5): 705–18.

Bagwell, Kyle. 2007. "The Economic Analysis of Advertising." In *Handbook of Industrial Organization*, vol. 3, edited by Mark Armstrong and Robert Porter. Amsterdam: Elsevier.

Bayer, Patrick, Fernando Ferreira, and Robert McMillan. 2007. "A Unified Framework for Measuring Preferences for Schools and Neighborhoods." *J.P.E.* 115 (4): 588–638.

Berndt, Ernst R. 1991. *The Practice of Econometrics: Classic and Contemporary*. Reading, MA: Addison-Wesley.

Berndt, Ernst R., Linda Bui, David R. Reiley, and Glen L. Urban. 1995. "Information, Marketing, and Pricing in the U.S. Antiulcer Drug Market." *A.E.R.* 85 (2): 100–105.

———. 1997. "The Roles of Price, Quality, and Marketing in the Growth and Composition of the Antiulcer Drug Market." In *The Economics of New Goods*, edited by Timothy F. Bresnahan and Robert J. Gordon. NBER Studies in Income and Wealth, vol. 58. Chicago: Univ. Chicago Press.

Berry, Steven T. 1994. "Estimating Discrete-Choice Models of Product Differentiation." *RAND J. Econ.* 25 (2): 242–62.

Berry, Steven T., James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica* 63:841–90.

Black, Sandra E. 1999. "Do Better Schools Matter? Parental Valuation of Elementary Education." *Q.J.E.* 114 (2): 577–99.

Busch, Susan H., Ezra Golberstein, and Ellen Meara. 2014. "The FDA and ABCs: The Unintended Consequences of Antidepressant Warnings on Human Capital." *J. Human Resources* 49 (3): 540–71.

Card, David, and Alan Krueger. 1994. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania." *A.E.R.* 84 (4): 772–93.

Ching, Andrew. 2010. "A Dynamic Oligopoly Structural Model for the Prescription Drug Market after Patent Expiration." *Internat. Econ. Rev.* 51 (4): 1175–1207.

Ching, Andrew, Robert Clark, Ignatius Horstmann, and Hyunwoo Lim. 2016. "The Effects of Publicity on Demand: The Case of Anti-Cholesterol Drugs." *Marketing Sci.* 35 (1): 158–81.

Ching, Andrew, Tülin Erdem, and Michael P. Keane. 2009. "The Price Consideration Model of Brand Choice." *J. Appl. Econometrics* 24 (3): 393–420.

Ching, Andrew, and Masakazu Ishihara. 2012. "Measuring the Informative and Persuasive Roles of Detailing on Prescribing Decisions." *Management Sci.* 58 (7): 1374–87.

Crawford, G. S., and M. Shum. 2005. "Uncertainty and Learning in Pharmaceutical Demand." *Econometrica* 73 (4): 1137–73.

Dickstein, Michael J. 2014. "Efficient Provision of Experience Goods: Evidence from Antidepressant Choice." Working paper, Stanford Univ.

Donohue, Julie M., and Ernst R. Berndt. 2004. "Effects of Direct-to-Consumer Advertising on Medication Choice: The Case of Antidepressants." *J. Public Policy and Marketing* 23 (2): 115–27.

Dube, Arindrajit, T. William Lester, and Michael Reich. 2010. "Minimum Wage Effects across State Borders: Estimates Using Contiguous Counties." *Rev. Econ. and Statis.* 92 (4): 945–64.

Dubé, Jean-Pierre, Günter J. Hitsch, and Puneet Manchanda. 2005. "An Empirical Model of Advertising Dynamics." *Quantitative Marketing and Econ.* 3 (2): 107–44.

Duggan, Mark, and Fiona M. Scott-Morton. 2006. "The Distortionary Effects of Government Procurement: Evidence from Medicaid Prescription Drug Purchasing." *Q.J.E.* 121 (1): 1–30.

Ellison, Glenn, and Sara Fisher Ellison. 2011. "Strategic Entry Deterrence and the Behavior of Pharmaceutical Incumbents Prior to Patent Expiration." *American Econ. J.: Microeconomics* 3 (1): 1–36.

Filson, Darren. 2012. "A Markov-Perfect Equilibrium Model of the Impacts of Price Controls on the Performance of the Pharmaceutical Industry." *RAND J. Econ.* 43 (1): 110–38.

Goeree, Michelle Sovinsky. 2008. "Limited Information and Advertising in the U.S. Personal Computer Industry." *Econometrica* 76 (5): 1017–74.

Hartmann, Wesley R., and Daniel Klapper. 2017. "Super Bowl Ads." *Marketing Sci.* Electronically published October 5.

Holmes, Thomas J. 1998. "The Effect of State Policies on the Location of Manufacturing: Evidence from State Borders." *J.P.E.* 106 (4): 667–705.

Iizuka, Toshiaki. 2004. "What Explains the Use of Direct-to-Consumer Advertising of Prescription Drugs?" *J. Indus. Econ.* 52 (3): 349–79.

Iizuka, Toshiaki, and Ginger Zhe Jin. 2005. "The Effect of Prescription Drug Advertising on Doctor Visits." *J. Econ. and Management Strategy* 14 (3): 701–27.

———. 2007. "Direct to Consumer Advertising and Prescription Choice." *J. Indus. Econ.* 55 (4): 771.

Ito, Koichiro. 2014. "Do Consumers Respond to Marginal or Average Price? Evidence from Nonlinear Electricity Pricing." *A.E.R.* 104 (2): 537–63.

Jayawardhana, Jayani. 2013. "Direct-to-Consumer Advertising and Consumer Welfare." *Internat. J. Indus. Org.* 31 (2): 164–80.

Kravitz, Richard L., Ronald M. Epstein, Mitchell D. Feldman, et al. 2005. "Influence of Patients' Requests for Direct-to-Consumer Advertised Antidepressants: A Randomized Controlled Trial." *J. American Medical Assoc.* 293 (16): 1995–2002.

Lewis, Randall A., and Dan Nguyen. 2012. "Wasn't That Ad for an iPad? Display Advertising's Impact on Advertiser- and Competitor-Branded Search." Working paper, Yahoo Res., Sunnyvale, CA.

Liu, Qiang, and Sachin Gupta. 2011. "The Impact of Direct-to-Consumer Advertising of Prescription Drugs on Physician Visits and Drug Requests: Empirical Findings and Public Policy Implications." *Internat. J. Res. Marketing* 28 (3): 205–17.

Liu, Qiang, Sachin Gupta, Sriram Venkataraman, and Hongju Liu. 2016. "An Empirical Model of Drug Detailing: Dynamic Competition and Policy Implications." *Management Sci.* 62 (8): 2304–20.

Liu, Qiang, Thomas J. Steenburgh, and Sachin Gupta. 2015. "The Cross Attributes Flexible Substitution Logit: Uncovering Category Expansion and Share Impacts of Marketing Instruments." *Marketing Sci.* 34 (1): 144–59.

Manchanda, Puneet, Peter E. Rossi, and Pradeep K. Chintagunta. 2004. "Response Modeling with Nonrandom Marketing-Mix Variables." *J. Marketing Res.* 41 (4): 467–78.

Murphy, Michael J., Ronald L. Cowan, and Lloyd I. Sederer. 2009. *Blueprints: Psychiatry.* 5th ed. Philadelphia: Lippincott Williams & Wilkins.

Narayanan, Sridhar, Ramarao Desiraju, and Pradeep K. Chintagunta. 2004. "Return on Investment Implications for Pharmaceutical Promotional Expenditures: The Role of Marketing-Mix Interactions." *J. Marketing* 68 (4): 90–105.

Nickell, Stephen. 1981. "Biases in Dynamic Models with Fixed Effects." *Econometrica* 49 (6): 1417–26.

Rosenthal, Meredith B., Ernst R. Berndt, Julie M. Donohue, Arnold M. Epstein, and Richard G. Frank. 2003. "Demand Effects of Recent Changes in Prescription Drug Promotion." In *Frontiers in Health Policy Research*, vol. 6, edited by David M. Cutler and Alan M. Garber. Cambridge, MA: MIT Press (for NBER).

Sahni, Navdeep S. 2016. "Advertising Spillovers: Evidence from Online Field-Experiments and Implications for Returns on Advertising." *J. Marketing Res.* 53 (4): 459–78.

Sinkinson, Michael, and Amanda Starc. Forthcoming. "Ask Your Doctor? Direct-to-Consumer Advertising of Pharmaceuticals." *Rev. Econ. Studies.*

Stremersch, Stefan, Vardit Landsman, and Sriram Venkataraman. 2013. "The Relationship between DTCA, Drug Requests, and Prescriptions: Uncovering Variation in Specialty and Space." *Marketing Sci.* 32 (1): 89–110.

Wittink, Dick R. 2002. "Analysis of ROI for Pharmaceutical Promotion (ARPP)." Assoc. Medical Pubs. http://kurse.fh-regensburg.de/kurs_20/kursdateien/2010Analysis_of_ROI.pdf.

Wosinska, Marta. 2002. "Just What the Patient Ordered? Direct-to-Consumer Advertising and the Demand for Pharmaceutical Products." Marketing Research Working Paper no. 02-04, Harvard Bus. School.

———. 2005. "Direct-to-Consumer Advertising and Drug Therapy Compliance." *J. Marketing Res.* 42 (3): 323–32.