

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

Bradley T. Shapiro*

This Version May 19, 2015

Abstract

Television advertising of prescription drugs is controversial, and it remains illegal in all but two countries. Much of the opposition stems from concerns that advertising directly to consumers may inefficiently distort prescribing patterns toward the advertised product. Despite the controversy surrounding the practice, its effects are not well understood. Exploiting a 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 then construct and estimate a multi-stage demand model that allows advertising to be pure category expansion, pure business stealing or some of each. Estimated parameters indicate that advertising has strong market level demand effects that tend to dominate business stealing effects. Spillovers are both large and persistent. Consistent with these spillovers, I find that firms advertise less and are less likely to advertise in markets with positive shocks to rival advertising. Using the demand estimates and a stylized supply model, I explore the consequences of the positive spillovers on firm advertising choice. Compared with a competitive benchmark, simulations suggest that a category-wide co-operative advertising scenario would produce four times as much advertising, resulting in a 18 percent increase in category size and a 14 percent increase in category profits.

1 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. 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 different product from the one advertised. This may be

*University of Chicago Booth School of Business. I thank Nancy Rose, Ernst Berndt, and Stephen Ryan for all of their advice. I would also like to extend my gratitude to Andrew Ching, Sara Fisher Ellison, Glenn Ellison, Günter Hitsch, Greg Leiserson, Sarah Moshary, Jesse Shapiro, Xiao Yu May Wang, and two anonymous referees for their helpful suggestions. This paper also benefited from conversations with Paul Snyderman and Len Tacconi; as well as participants at the MIT Industrial Organization seminar and field lunch. 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 Number T32-AG000186 and the National Science Foundation Graduate Research Fellowship under Grant Number 1122374. All mistakes are my own.

due to difficulty in remembering advertisements, agency problems in obtaining products or simply because 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 this, I construct and estimate a demand model which allows such spillovers. Given these spillovers, I test to see if firms free ride off of rivals' advertising. 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 jointly decide advertising, and I compare this outcome to a benchmark competitive outcome in the antidepressant market.

Branded television advertising of prescription drugs is contentious and has been condemned by many as inefficiently distorting prescriptions to the advertised products. In fact, it is legal in only two countries: New Zealand and the United States. In light of the controversy, it is important to understand the impact of these advertisements. 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. This could be either good or bad for social welfare depending on whether or not 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. While cooperation is often difficult to enforce and non-contractible due to antitrust laws, advertising cooperatives are precedented in other industries such as 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 (Dube et. al. 2005). However, such specifications also typically exclude the possibility of positive spillovers of advertising onto rivals. While this eliminates the complexity of modeling behavior in the presence of possible free riding, such an exclusion 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 which facilitate positive spillovers in television advertising. First, the FDA regulates what firms can and cannot say in advertisements. While 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 there are several therapeutic products available, those treating the same ailments tend to share common characteristics. A consumer might remember all of 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 excluded by typical discrete choice specifications. Positive spillovers are allowed, but not imposed by the model. Further, I find that not using the border discontinuity and assuming advertising choices are exogenous leads the researcher to over-state the long run effectiveness of advertising as well as understate the extent of the positive spillover. Next, I test whether firms free ride off of rival advertising efforts. I find that firms advertise less and less often when there are positive shocks to rival advertising 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 with a scenario whereby a co-operative sets advertising for the entire industry.

While most models incorporating advertising into demand have not allowed for positive spillovers, there are studies of Direct-to-Consumer (DTC) advertising in pharmaceuticals with varying credibility of identification strategies that 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, [Iizuka & Jin (2005), Berndt et. al. (2004), Wosinska (2002)], find very small estimates of advertising effects on market shares conditional on being in the market and conclude that it might be exhibiting positive spillovers, though spillovers are neither directly modeled nor tested. Wosinska (2005) and Donohue et. al. (2004) find that advertising has positive spillover effects onto drug compliance and duration of treatment. Other studies find that advertising drives consumers to the doctor (Iizuka & Jin 2004) or has class level effects (Rosenthal et. al. 2003, Avery et. al. 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 on brand shares, focusing mostly on physician detail advertising and academic journal advertising since DTC was extremely limited and unbranded at the time, and found some effects at both category and product levels. In studying detailing effects, Ching et. al. (2012) take advantage of the fact that identical molecules are sometimes marketed by different firms under different names in Canada to separate out brand versus category effects. Narayanan et. al. (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 only affect class level demand and finds significant effects. Many of these studies either only model a category level response or only model a conditional share level response. This paper will model the full decision process and use data with both spatial and time series variation. Stremersch et. al. (2013) and Liu & Gupta (2011) also examine the various effects of DTC on aspects of demand¹. Stremersch et. al. can explain variation across geography using demographic characteristics and both find heterogeneous effects. This study will differ from both of those in that fixed effects will be used 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, there might well be an incentive to invest less in advertising. Iizuka (2004) finds that as the number of competitors increases, firms advertise less, leading him 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

¹Stremersch et. al. (2013) looks at the effects of DTC through the mediator of patient requests and finds no effects. Liu & Gupta (2011) use information on patient visits.

the magnitude of the potential positive spillover effects on advertising expenditure decisions. Ching (2010), Filson (2012), Liu et. al. (2013) all use a Markov Perfect Equilibrium concept to model the supply side of pharmaceutical markets, but they all focus on different aspects of the pharmaceutical industry rather than television advertising.

Outside of the pharmaceutical literature, Sahni (2013) finds experimental evidence of positive spillovers to rivals in online restaurant advertising in India. Additionally, Lewis & Nguyen (2012) and Anderson & Simester (2013) find evidence of positive spillovers in a number of categories for online and mail advertising, respectively. Non-experimentally, Ching et. al. (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.

The contributions of this paper are threefold. First, I improve upon 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 due to 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. I construct and estimate a consumer choice model, which 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 carry-over 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 over-states 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 over-state the long run effectiveness of advertising and under-state the relative long run importance of category expansion relative to business stealing. Next, as the estimated demand parameters imply an incentive for firms to free ride, I directly test whether firms advertise less or less often in markets where rivals have idiosyncratically high advertising. Consistent with these incentives, I find that firms advertise less and less often in markets where rival advertising is high. Finally, I conduct a supply side analysis using a stylized model to evaluate to what extent positive spillovers suppress the incentive to advertise. Given the demand parameters, I compute a benchmark competitive outcome in television advertising. I find that if instead, firms that advertise work together, removing the need for strategic response, those firms would combine to advertise 50% more than in competitive equilibrium. A co-operative deciding all advertising expenditure levels taking full industry profits into account would advertise four times as much as is observed in competitive equilibrium, increase the category size by 18% and category profits by 14%. 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 who provide it, and understanding both are essential to firm profit maximization and to efficient regulation.

2 Empirical Setting

2.1 Prescription Drugs and Advertising

Television advertising of prescription drugs did not appear in the United States until 1997. While technically not forbidden by law, advertising was required to have much more risk information included on all advertisements 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 thirty second spot was prohibitively time consuming and costly. In the fall of 1997, the FDA issued a draft memorandum clarifying their stance on advertising risk information, allowing advertisements to air so 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 pre-approval to ensure that the ‘fair balance’ condition was met. In 1999 the final copy of the FDA memorandum was circulated. The first advertisements on television for antidepressants were seen 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.

2.1.1 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 those 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 allow patients to be more safely treated and with fewer side effects (Anderson 2000). This allows easier management of antidepressant treatment by primary care physicians, and makes 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 have 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 that antidepressants might lead to an increase in suicidality among adolescents (Busch et. al 2012). 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, ten other brands, some with slightly different mechanisms, have been discovered and have entered the market. Some of those have developed extended release versions which allow patients to have fewer doses per day.

There are six main subcategories of antidepressants: the old style TCAs, Tetracyclic (TeCA), Serotonin Antagonist and Reuptake Inhibitors (SARI), Serotonin-norepinephrine Reuptake Inhibitors (SNRI), Norepinephrine Reuptake Inhibitors (NDRI) and SSRI. While the specific differences between these are not important to this study, it is worth noting 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 is included on the patient’s insurance formulary or physician preferences. 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 et. al. 2009).

Many physicians see depression as an under treated condition and some research has concluded that restricting access to antidepressants has been associated with negative health outcomes (Busch et. al. 2012). Given this information, it is plausible that market expansive advertising could play a role in this market.

2.1.2 The Market for Advertising

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

Additionally, there are both national and local advertisements. National advertisements are seen by everyone in the country tuned into a particular station, while local advertisements are only seen by households within a particular designated market area (DMA).

A DMA is a collection of counties, typically centered around a major city, and it 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 that 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 times was just 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 FCC moderates the signals to try to keep the signal from each station localized only in its own DMA.³ There are 210 DMAs in the United States, the largest 101 of which are included in my data.⁴

From informal conversations with individuals in industry, I learned that pharmaceutical companies participate almost exclusively in the up front market. Like 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.

Prices for advertisements typically are determined by projected volume and type of viewership. A single airing of a national advertisement for antidepressants 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

²<http://www.sbca.com/dish-satellite/dma-tv.htm>

³<http://www.fcc.gov/encyclopedia/evolution-cable-television>

⁴I note here that from time to time, Nielsen may move one county from one DMA into another. In this data, I have a snapshot of DMA composition currently. Discussions with Nielsen have assured me that these shifts are sufficiently infrequent and few that using current DMA information should not be problematic. To the extent that a county gets categorized in the wrong DMA, it will lead to measurement error that will bias estimates towards zero.

advertisements range from \$0.0002 per 100 to \$0.04 per 100 and 93% 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 at all. 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 among all of the pharmaceutical industry (Berndt et. al. 2003). The lack of advertising from all firms could be indicative of fixed costs of advertising at all or of free riding. Those branded products which do not advertise either have low market share (Effexor XR, Remeron, Serzone) or have a very small parent company which might be less likely to have an advertising division (Celexa, Lexapro). Whether or not we can observe free riding will be further evaluated in the supply analysis.

2.2 Data

2.2.1 Prescribing Data

Sales data for this market comes from the Xponent data set of IMS Health, a health care market research company. The prescribing behavior of a 5% random sample of physicians who prescribe antidepressants is followed monthly from 1997 until 2004. The data include a rich set of physician characteristics including address of the primary practice, which is then linked to county. The data used in this study is aggregated to the county level and ends with 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.

2.2.2 Advertising Data

Product level monthly advertising data at the national and Designated Market Area (DMA) level for the top 101 DMAs comes from Kantar Media. In addition to advertising expenditures, the data includes 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 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 level for the period of the data where 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 as well as national 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 there is non-trivial variation 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 a great deal.

⁵A possible alternative measure would be to use 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, as a commercial during the evening news is likely to capture far more eyeballs than a commercial during a 1:00 AM rerun of MacGyver, using expenditures per 100 capita would seem to do a better job at measuring quality adjusted advertising than number of commercials.

2.2.3 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 2134 general practice physicians are followed monthly through time from 2001 through the end of 2003, and a panel of 167 psychiatrists are followed monthly for 2002 and 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 non-representativeness is due to the fact that these are the physicians who are most likely to ever be detailed. While these physicians make up less than 1% of total physicians in the country, they are likely to make up a significantly higher percentage of both the prescription and detailing distributions. Additionally, national aggregate detailing data by brand from IMS Health is observed in the data.

2.2.4 Other Data Sources

I observe prices from Medicaid reimbursement data, collected by the Centers for Medicare & Medicaid Services (CMS). Duggan and Scott-Morton (2006) argue that the average price that 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). As I will not be estimating marginal production costs empirically, these average pharmacy acquisition costs may be used as an effective upper bound on marginal production costs. While there are markups from branded drugs, pharmacies are typically able to obtain generics at much lower rates, particularly when there are several generic competitors (as is the case in this market), often as low as ten 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 version 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 (CPS).

3 Reduced Form Evidence

In this section, I explore the data to see if spillover effects exist and how they interact with own effects. This exercise has been difficult to implement in previous research, largely due to data limitations. Estimates show that rivals' and own advertising have a positive effect on sales, while 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 due to decreasing returns to scale.

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

$$\log(Q_{jmt}) = \lambda \log(Q_{jmt-1}) + \gamma_1 a_{jmt}^{own} + \gamma_2 a_{jmt}^{cross} + \gamma_3 (a_{jmt}^{own})^2 + \gamma_4 (a_{jmt}^{cross})^2 + \gamma_5 a_{jmt}^{own} a_{jmt}^{cross} + \varepsilon_{jmt} \quad (1)$$

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

3.1 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 which 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 et. al. (2012) 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 et. al. (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 very little correspondence with anything else in the world. As such, we might think the location of DMA borders is far more exogenous to consumer characteristics than are state borders.

Advertising is purchased both nationally and locally. The level of total advertising that a household gets to its television is determined by the Designated Market Area (DMA) that the household's county belongs to, as defined by AC Nielsen. Nielsen places counties into markets by predicting which local stations the households will be most interested in. As such, DMAs tend to be centered at 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 mustard color Cleveland, Ohio DMA gets the same amount of the same advertising as every other county in the Cleveland DMA. Meanwhile, every county in the green color Columbus, Ohio, DMA gets the same amount of the same advertising as every other county in the Columbus DMA, though this might be different from the advertising in the Cleveland DMA. Meanwhile, these two DMAs border each other. There are five counties in the Cleveland DMA which share a border with at least one county in the Columbus DMA and five counties in the Columbus DMA which share at least one border with a county in the Cleveland DMA. My strategy will be to consider these ten 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. This means that a group of counties along a particular border but

in the same DMA are aggregated together, as they see 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. For each border experiment in each time period, there will be 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 only enforced locally at the border between two DMAs, allowing for spatial heterogeneity. Border-DMA fixed effects will allow systematically different demand levels across 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 being treatment groups. The fixed effects specification is:

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

where the subscripts j and b indicate which experiment is being considered (product and border specific), α_{jbq} is a time effect which 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 $f(a_{jmt})$ is the magnitude of the treatment. The magnitude of the treatment is zero everywhere prior to 1999, as the FDA memo had not yet gone into effect. To investigate persistence in demand, a lagged dependent variable is also included. It should be noted here that the inclusion of α_{jbm} in the specification means that I am focusing on market level deviations from trend. That is, each market has a fixed effect. While Stremersch et. al. (2013) find that the distribution of DTC across markets may be explained by region-specific demographic composition, 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 Zolofit advertisements. In the equation above, $\log(Q_{jbmt})$ is log number of prescriptions of Zolofit in the Cleveland-Columbus border, indexed by month and which side of the border it is on. The magnitude of the treatment, $f(a_{jmt})$ is a function of the Zolofit’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 absent advertising differences. Is this plausible? These counties are bordering, so they are very 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, as 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, there will be at least some variation in each experiment over time.

While it is clear that advertising is a firm choice rather than completely random, it is instructive to think about the potential sources of endogeneity and how the border strategy addresses those specific sources. Since there are market level fixed effects, endogeneity that comes from, say, the fact that winters in Florida are milder than winters in Wisconsin is not a concern. Those types of concerns are absorbed in the market level fixed effects. Potential bias can only come 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. Since 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 there is no reason that the weather should 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 very 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 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 further away a person is from a place of employment, the less likely he or she will work there, due to 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 might increase the use of some antidepressant while also increasing advertising to the DMA where it occurs. Since these seminars also tend to be in center of DMAs, those at the outskirts are less likely to attend due to transportation costs, but those transportation costs do not discontinuously change at the DMA border. These three unobserved shocks seem to be the most likely to drive advertising decisions at the DMA level, and all are addressed by the border approach. It is important to note that there might be other potential unobserved shocks. However, so long as these shocks are reasonably continuous in distance from the center of the DMA and do not discontinuously change at the border, the border approach will be valid. All of these requirements hold for the aforementioned examples.⁶

Next, consider the possibility that firms use rules of thumb to allocate advertising based on the demand in the previous period. If this is the case, advertising in a DMA in the current period is determined by some function of last period's

⁶While most DMAs are centered around a large city, a few DMAs have two main cities (e.g. Johnstown-Altoona). If the main targets of a DMA are at the borders, for the border strategy to remain valid, we need only that the demand shocks immediately across the border are the same. If the firm targets those specific demand shocks, it is likely that across the border, there will be no variation, as 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, as we are controlling for the demand shocks directly using the borders.

demand across the DMA. If previous period demand is correlated with current period demand, estimates will be biased. This 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 comprise a fraction of the population, sales and counties in a DMA. Even if the trends for demand are exactly 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 variables 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 get soaked into the product-border-time fixed effect, as demand is very 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 is a problem in this setting since 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.

3.1.1 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. It might be the case that the true treatment effect in the interior of the DMA is different from that at the border. If this is the case, the interpretation of the supply analysis will be limited as I assume that the estimated advertising effects hold in both border and non-border counties. This concern is partially addressed in a robustness check in the appendix. 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 further out. The effects are very similar and not statistically distinguishable from the full border sample. While this 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 those 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 non-federal physicians. To the extent that we continue to worry that the estimated treatment effects are different at the interiors, that 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.

3.1.2 Potential Threats to the Border Strategy

One potential worry is that there would be little variation net of the fixed effects. This would be the case if too much of the advertising were national and not enough were 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 there is substantial variation net of fixed effects.

Also potentially problematic is the lagged dependent variable, which can generate omitted variables bias in the presence of small T, as differencing mechanically induces correlation between the lagged dependent variable and the error term. However, as $T \rightarrow \infty$, the mechanical correlation with the error term diminishes to zero and the fixed effects estimator is consistent. As my data is 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. There are a three main possibilities that 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 towards 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 under-state the effect of the advertisements. To the extent that we think that these biases are present, we view these estimates as lower bounds on the true parameters.⁸

It should also be noted that very few consumers watch over the air. According to the Consumer Electronics Association, fewer than seven 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, as stations tend not to locate in the outskirts of DMAs. It is likely that consumers would be even less likely to rely on over the air signals at the DMA border.

Omitted variables bias could also be a source of bias. Prices, magazine and newspaper advertising and detailing are omitted from this estimation. As prices tend not to vary geographically due to 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 and will be absorbed into the fixed effects, as they only vary over time. Any national average effects of detailing are also controlled for by the product-border-time effect.

However, if firms strategically raise (or lower) detailing at the product-market-time level in exactly the same places where DTC is concentrated to take advantage of any complementarities or substitutabilities, it will induce bias. For a monthly panel of 2134 general practitioners and 167 psychiatrists in 2001-2003, ImpactRx data has physician specific

⁷In Nickell (1981)'s paper that 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: $\lim_{N \rightarrow \infty} (\hat{\lambda} - \lambda) = \left\{ \frac{2\lambda}{1-\lambda^2} - \left[\frac{1+\lambda}{T-1} \left(1 - \frac{1-\lambda^T}{1-\lambda} \right) \right]^{-1} \right\}^{-1} \approx \frac{-(1+\lambda)}{T-1}$, where the approximation holds for "reasonably large" T. With $T=84$, that approximation is bounded above by $-\frac{2}{83} \approx -0.02$. 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. As this only holds as $N \rightarrow \infty$, I ran simulations assuming the data generating process is as estimated and found the magnitudes to be nearly identical. Details of the simulation are available from the author upon request.

⁸However, the Dartmouth Institute has drawn primary care commuting zones which describe how far Medicare patients travel to see their physicians. It is very rare for a commuting zone to cross DMA lines- only about 1% of primary care commuting zones cross DMA borders at all, and those that do tend to be predominantly in only one DMA. This should minimize the measurement error worry. Further explanation of the Dartmouth Institute commuting zones is provided in the appendix.

⁹<http://www.tvtechnology.com/default.aspx?tabid=204&entryid=9940>

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 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 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 variables 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 (2001-2003) and markets for which there is detailing data and show that the inclusion of detailing does not affect estimates of the effect of DTC. Since the number of time periods are greatly reduced when using this 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 there is no trend break in firm detailing strategy nationally at that point where DTC changes drastically from zero to significantly positive.

This lack of coordination may seem surprising on the face of it, as detailing and DTC advertising are two very important pieces of the marketing mix for pharmaceutical firms. There are strong institutional reasons why firms may fail to coordinate these efforts during the sample period. In particular, the organization of the firm and the nature of sales representative employment make coordination very difficult. Those managers who decide DTC advertising tend to be in the consumer division of the firm while 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. These deciles rules are documented in the literature (Manchanda et. al. 2004). The basic intuition is that the analytics company groups physicians into ten decile buckets based on prescription volume. They then suggest that sales representatives visit each physician in the same decile the same amount. The deciles are adjusted over time as physicians change their intensity of prescribing, but adjustments more frequent than yearly tend to be very 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 there to be systematic coordination of DTC with detailing, the sales reps would either all have to decide to work harder during high DTC months or more sales reps would have to be temporarily hired. Neither of these is easily executed, particularly on a month-to-month basis.

A further concern is that various policies or cost inputs could discontinuously change right at the DMA border, causing the more impressionable physicians to locate on a particular side of the DMA border. As DMAs are in general only relevant to television markets, it is hard to imagine why any tax laws would systematically vary across DMA borders. Almost all business tax policies are set by state governments or potentially large city governments. Being on the border

¹⁰Conversations with experienced sales representatives confirm this intuition.

of the DMA typically leaves those counties out of reach of large city specific taxes. However, as 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 re-estimated 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.

Data on corporate rental rates by county is not available for this study, so it is impossible to assess whether or not there is a discontinuity in the rental rates of physician offices across a DMA border. However, it is hard to imagine that within a state for two very similar counties which border each other and are similar distances from major cities that rental rates would differ significantly. Furthermore, if they did, it is hard to imagine that advertising decisions for the full DMA would hinge on rental rates at the border of the DMAs. The only worry is if the physicians that select into cheaper rent are systematically those which have different responsiveness to advertising than those who select into more expensive rent. To further examine the issue of selection across the border, data was collected from the Area Resource File to see if the number of physicians, the average income or the population is significantly different on the higher advertising side of the border from 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. It could be that the difference-in-differences model fails 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.

3.1.3 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 BLP 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 use of the BLP instruments would eliminate any spatial variation, which is a main attribute of the data I am using. Furthermore, it might be unreasonable to think that competitor characteristics are exogenous in this setting. It stands to reason that as consumers demand more antidepressants with fewer of some kind of side effects that firms might well focus research and development on that kind of product.

3.2 Results

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

$$\log(Q_{jmt}) = \lambda \log(Q_{jmt-1}) + \gamma_1 a_{jmt}^{own} + \gamma_2 a_{jmt}^{cross} + \gamma_3 (a_{jmt}^{own})^2 + \gamma_4 (a_{jmt}^{cross})^2 + \gamma_5 a_{jmt}^{own} a_{jmt}^{cross} + \alpha_{jbq} + \alpha_{jbd} + \varepsilon_{jmt} \quad (2)$$

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 Wisconsin).

Partialing these fixed effects out 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 that either side of border b .

Results of the above regression are provided in Table 2. Most notable is that both rivals' and own advertising has 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 exactly where the incentive to free ride comes from, as rival advertising lowers the marginal revenue of own advertising. Finally, there is evidence of persistence, though the persistence parameter is not especially large. This is consistent with the idea that there is much experimentation to find the correct fit between patient and treatment in the depression space.

4 Model

4.1 Demand

I propose a multi-stage choice model where 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, the consumer 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 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). This basic structure of this demand model is similar to Berndt et. al. (1997) and Ching et. al. (2011).

I define 'utility' u of consuming the inside option, as a function of total advertising stock as well as other market level factors:

$$u_{ilm} = \Gamma_1(A_{ilm}) + \beta_1 X_{ilm} + \alpha_t + \alpha_m + \xi_{ilm} + \varepsilon_{ilm} = \delta_t + \varepsilon_{ilm}. \quad (3)$$

In this specification, l denotes the inside option versus outside option, m denotes market and t denotes time period. I define Γ_1 as an increasing function of A_{ilm} , total advertising stock of all inside option products in market m at time

t , α_{it} is a time specific taste for the inside option, α_{im} is a market specific taste for the inside option, and X_{imt} are market-time characteristics.

For the next stage, I define the relative utility v^l of subcategory n conditional upon 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:

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

Finally, relative utility w^n of product j conditional upon 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:

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

Dynamics enter the model through advertising carry-over. 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 .

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

I set each disturbance term, ε , to be iid 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:

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

Those conditional shares take logit form:

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

$$s_{n|l} = \frac{\exp(\delta_{n|l})}{1 + \sum_n \exp(\delta_{n|l})} \quad (9)$$

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

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 while I call the latent variables at each level ‘utilities’, it is not essential to interpret them literally as such. In this paper, I will not be computing consumer welfare, and it is likely that the latent variables contain a combination of patient and physician utility, information and persuasion. The purpose of the choice model is to guide the firm decision problem. While it is possible that these parameters could 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, as discovering the relative magnitudes of advertising effects at each level is a main question of this paper.

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, there might be different effects 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, the percent black, percent Hispanic, percent Asian, income, percent uninsured, percent over age 45 and the employment to population ratio are used.¹² Including heterogeneity over and above these fixed effects and demographic interactions resulted in no significant findings, perhaps because so much variation is explained by these fixed effects. Descriptive statistics for the demographic variables are available in Table 4. For ease of interpretation, when included in the demand model, these demographics are log-normalized with mean zero and standard deviation of one.

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 then in turn increases $\Gamma_1(A_{lmt})$. Next, the information in the advertisement may push the consumer towards the subcategory of antidepressants that Zoloft is in over another, as 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 done 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.

It is also worth noting that the model does not explicitly examine the various possibilities for forward looking consumers, including consumer learning, as in Dickstein (2014), Ching (2010) or Crawford & Shum (2005). However, the difference in persistence parameters from the category level to the product levels allows for consumers to purchase one brand, decide that 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, it suggests the consumer is still consuming in the category but no longer with the same product. Further, the product-time fixed effects allow differences in market conditions to be taken into account. That is, in the year prior to Prozac going off patent, consumers knowing that 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.

¹¹As 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.” It is hard to imagine why a relative preference between Prozac and Zoloft should affect a consumer’s absolute taste for antidepressants.

¹²These are the same demographic interactions as are included by Stremersch (2013) plus percent uninsured and employment to population ratio.

4.1.1 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:

$$\frac{\partial s_j}{\partial a_k} = s_{j|n} [s_{n|l} \frac{\partial s_l}{\partial a_k} + s_l \frac{\partial s_{n|l}}{\partial a_k}] + s_{n|l} s_l \frac{\partial s_{j|n}}{\partial a_k} \quad (11)$$

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

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

and advertising elasticities equal to,

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

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 $s_j \frac{\partial \Gamma_1}{\partial a_j} (1 - s_l)$, through expansion of the subcategory in $s_j \frac{\partial \Gamma_2}{\partial a_j} (1 - s_{n|l})$ and through business stealing within the nest in $s_j \frac{\partial \Gamma_3}{\partial a_j} (1 - s_{j|n})$. Firm benefits from rivals' advertising in the same subcategory may flow through expansion of the category in $s_j \frac{\partial \Gamma_1}{\partial a_k} (1 - s_l)$ or through expansion of the subcategory in $s_j \frac{\partial \Gamma_2}{\partial a_k} (1 - s_{n|l})$, while this same advertising may hurt through business stealing within the subcategory in $-s_j \frac{\partial \Gamma_3}{\partial a_k} s_{k|n}$. 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. It is worth noting that this structure fully allows for advertising that is a pure category expansion (i.e. if $\frac{\partial \Gamma_2}{\partial a_j} = \frac{\partial \Gamma_3}{\partial a_j} = 0 \forall j$), for advertising that is pure business stealing (i.e. if $\frac{\partial \Gamma_2}{\partial a_j} = \frac{\partial \Gamma_1}{\partial a_j} = 0 \forall j$), or anything in between, including cross subcategory substitution. It is also possible that rival advertising outside of the subcategory could help more than inside of the subcategory if $\frac{\partial \Gamma_2}{\partial a_j}$ is sufficiently small and $\frac{\partial \Gamma_3}{\partial a_j}$ 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.

Notable is 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 there are decreasing returns 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, while the fact that the coefficient on the cross partial is negative leads to the second.

5 Empirical Specification and Estimation of the Model

5.1 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, similar to Dube et. al. (2005).

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

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

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

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

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 levels are shares and advertising.

Substituting equations (14) and (15) into equations (3)-(5), for the market level l obtain:

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

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

5.2 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. This results 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 :

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

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:

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

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:

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

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

$$\log(s_{lmt}) - \log(s_{omt}) = \lambda_l (\log(s_{lmt-1}) - \log(s_{omt-1})) + \gamma_1 \log(1 + a_{lmt}) + \theta_{lt} + \theta_{lm} + v_{lmt} \quad (20)$$

where

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

is a inside option-time specific taste or quality parameter.

and

$$\theta_{lm} = \alpha_{lm} - \lambda_l \alpha_{lm} \quad (22)$$

is the category-market specific taste parameter. Finally,

$$v_{lmt} = \xi_{lmt} - \lambda_l \xi_{lmt} \quad (23)$$

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:

$$\log(s_{nmt|t}) - \log(s_{0mt|t}) = \lambda_n(\log(s_{nm,t-1|t}) - \log(s_{0m,t-1|t})) + \gamma_2 \log(1 + a_{nmt}) + \theta_{nt} + \theta_{nm} + v_{nmt} \quad (24)$$

$$\log(s_{jmt|n}) - \log(s_{0mt|n}) = \lambda_p(\log(s_{jm,t-1|n}) - \log(s_{jm,t-1|n})) + \gamma_3 \log(1 + a_{jmt}) + \theta_{jt} + \theta_{jm} + v_{jmt}. \quad (25)$$

5.3 Identification and Estimation Strategy

Since the share equations have been transformed to a linear form, estimation may be done by OLS. 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 (3).

In particular, I specify the estimation equation as:

$$\log(s_{lmt}) - \log(s_{0mt}) = \lambda_l(\log(s_{lm,t-1}) - \log(s_{0m,t-1})) + \gamma_1 \log(1 + a_{lmt}) + \theta_{lbq} + \theta_{lbt} + v_{lmt} \quad (26)$$

where θ_{lbq} is a border-time fixed effect and θ_{lbt} 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 that 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_{nbt} and at the product level, I include fixed effects α_{jbq} and α_{jbt} . Identifying variation will come at the subcategory level from total subcategory advertising that is above and beyond 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 3 has variable definitions and summary statistics for those variables that will enter the estimation.

5.4 Demand Results

5.4.1 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 inside and outside option and at the product business stealing level. Effects at the subcategory level are not significant, but it is notable that there is only advertising in two subcategories, with most of the advertising happening in the SSRI subcategory. The small and insignificant effect

at the subcategory level is not surprising, as it seems unlikely that patients would have good information about what separates the subcategories. 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 female. All other specifications in the paper will include these demographic interactions, but will suppress them, as their inclusion or exclusion do not affect the estimated main effects.

Table 7 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.

5.4.2 Persistence

Persistence is highest at the category level. The persistence parameter of 0.68 implies that 90% of the effect dissipates within six months. Meanwhile, the 0.32 persistence parameter at the product level implies that 90% of the business stealing effect of an advertisement dissipates within only two months. This makes advertising in the long run more of a category expansion than a business stealing tool. This 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 quitting antidepressants altogether. It is also consistent with a limited memory view of advertising. Since advertising 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 co-operative.

5.4.3 Importance of Using the Border Strategy

Table 6 presents estimates that highlight the importance of using the border strategy to account for the endogeneity of firm choice. In the first column, the demand analysis is done at the DMA level without using market or time fixed effects. Market fixed effects control for persistent differences in demand patterns while the time fixed effects control for national market conditions, such as patent expirations, new clinical study results and new product introductions. The second column includes these fixed effects in a DMA level analysis, but does not control for the endogeneity of firm choices using the border strategy. The third column uses the border strategy and fixed effects.

Notable from these results is that failure to use fixed effects under-states the category level effects and over-states the business stealing effects of advertising. In addition, persistence parameters are drastically over-stated. 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 over-statement of the business stealing effect. Further, generic introductions likely skew advertising decisions. Branded products on the whole would prefer to advertise more in markets where there is 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 due to the lower priced generics. This leads the researcher to under-state the category

expansive effects of advertising. Both of these concerns may be mitigated by controlling for generic introductions and other market conditions using time fixed effects as well as controlling for persistent market differences using DMA fixed effects.

Upon introducing the DMA and time fixed effects in the second column, 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 over-stated. This 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 over-stated. Given that the over-statement of these persistence parameters is more severe at the subcategory and product levels, it will lead the researcher not only to over-state the long run effectiveness of advertising, but will also lead to an under-statement of the magnitude of the spillovers over time.

6 Supply and Counterfactual

6.1 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 which are not internalized by the advertising firm. Second, rival advertising lessens the incentive to advertise by lowering the marginal category expansive effect of advertising.

6.1.1 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, while 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. There is a large contemporaneous positive spillover that intensifies through time. Zoloft has no incentive to internalize the benefits it bestows upon other firms, and thus will under invest in advertising relative to a co-operative controlling advertising in the whole market.

6.1.2 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 that 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. Notable from the curves is 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, it is almost not worthwhile for Zoloft to advertise at all. There is a clear incentive for Zoloft to free ride as competitors advertise more and more.

6.1.3 Do Firms Actually Free Ride?

While the last two sections show that the demand estimates imply an incentive for firms to under invest through internalization failures and free riding, they do not provide evidence of actual firm free riding. To evaluate whether or not firms free ride, we can test whether or not 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:

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

where f is an increasing function of own product advertising and g is a function of rival product advertising. In Table 8, this question is addressed in four different ways. In the first two columns, 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 the first column, $f(a_j) = a_j$ and $g(\sum a_{-j}) = \sum a_{-j}$. 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 \$0.02. This is a small but significant free riding effect. In the second column, this is evaluated in logs, where $f(a_j) = \log(a_j)$ and $g(\sum a_{-j}) = \log(\sum a_{-j})$. When rivals increase advertising by 10%, firms decrease advertising by 0.2%. The second two columns address the extensive margin. Are firms less likely to advertise at all in a market if rivals advertise more? In the third column, $f(a_j) = \mathbf{1}(a_j > 0)$ and $g(\sum a_{-j}) = \sum a_{-j}$. As rivals increase DTC by \$1 per 100 capita, a firm's likelihood of advertising in that market decreases by 0.003. Finally, in the fourth column, $f(a_j) = \mathbf{1}(a_j > 0)$ and $g(\sum a_{-j}) = \mathbf{1}(\sum a_{-j} > 0)$. The presence of rival advertising in the market decreases a firm's likelihood of advertising in that market by about 0.068. In all specifications, rival advertising is negatively and statistically significantly associated both with the amount of firm advertising and the decision of whether or not to advertise, suggesting that firms have some understanding of the free riding incentives generated by spillovers in this market.

6.2 Supply Simulation

While the previous section shows that firms are free-riding, it does not speak to the size 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. This will be done by plugging in 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.

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 that firms in fact did not have this level of sophistication at the time of the sample. In 2002, the Association of Medical Publications (AMP) commissioned the Analysis of Return on investment for Pharmaceutical Promotion (ARPP) study (Wittink 2002). This study had a steering committee including representatives from Wyeth, 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 that 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 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.

The model will be based on a Markov Perfect Equilibrium (MPE) concept that only allows firms 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 firms have a monthly discount rate of 0.95 and a marginal cost of advertising \$1 of \$1.15, including both the pecuniary cost of advertising and a 15% agency fee. In this way, all inputs into the model are observable, and all that remains to do 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 co-operative 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 under investment incentives.

One complication of solving a 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 (that is, their market shares get very low and they cease to participate in the advertising game). In the demand estimation, these issues are dealt with using product-time fixed effects. In principle, we might view these effects as additional state variables. Since entry and exit are not the main focus of this study, my strategy will be to focus on a period within my sample where the market is stable. From December 1999 until December 2000, there are no product entries or exits. The next major product entry is July of 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, since the effects of advertising are almost entirely dissipated over the course of seven 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 of 2001.

For the period described, only two products advertise on television: Paxil and Prozac. Paxil is produced by Glaxo-SmithKline, 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, Georgia DMA. In particular, prior to the period in question, the only advertising done was by Paxil. That 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 there 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% of the population to about 6.7% of the population over the course of the year. Profit evolution for each firm in the benchmark scenario is provided in Figure 12. 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, 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 with what is observed in the data (shown in Figure 13), Prozac advertises far less than would be predicted from the demand model, suggesting that they are free riding more than the demand parameters would suggest that they should. This could be a result of organizational costs or budget constraints, or simply a matter of learning the game, as DTC was relatively new to the pharmaceutical industry during this period. Paxil sometimes advertises a little more and sometimes advertises a little less than the demand parameters would suggest that they should. As mentioned, the purpose of this model is not to predict the observed levels of advertising but to illustrate the magnitude of incentive effects. Given the existence of the ARPP study, it might not be surprising that these numbers do not match the data.

6.2.1 Counterfactual Scenarios

I use two alternative scenarios to size the incentive effects of positive spillovers. First, I assume that Eli Lilly and GlaxoSmithKline co-operate on advertising that maximizes the combined profits of their products, which takes away the need for strategic response. Next, I'll consider a scenario whereby the entire market is allowed to set advertising in a single optimization problem to think about the potential of a category-wide advertising co-operative. The ability to contract on coordination would allow firms to overcome the free riding problem and provide advertising, even without a brand level component.

For the purposes of the counterfactuals, I assume that the advertising firms in the antidepressant market cooperate to make a common non-branded category advertisement for antidepressants, facilitated by a patient advocacy group. Such groups are focused on educated 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 treatment 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. As none of these types of advertisements are observed in the data, this effect is assumed. We could alternatively assume that the co-operative agrees that only one product uses branded advertisements and then transfers the business-stealing rents via a pre-arranged contractual arrangement, and the analysis would follow in exactly the same manner.

The co-operative solves the firm's problem in each month and in each market. In the Eli Lilly/GSK scenario, the co-operative includes Paxil, Prozac, Wellbutrin and Wellbutrin SR as part of the portfolio. In the full co-operative scenario, all products in the category are included as part of the portfolio in the firm's problem. The marginal co-operative advertisement dollar has cost equal to \$1.15, as before. The co-operative solves the same optimization problem as in the benchmark, but since 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 co-operative advertising campaigns. This is for two main reasons. First, neither prices nor detailing are responsive to DTC in the data. This is shown empirically in Appendix A. The pricing result is also consistent with previous research which shows that 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. It must be noted that while detailing is not associated with DTC in the data, a co-operative 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. It is also worth noting that since in the benchmark scenario, firms under invest in advertising, the counterfactual scenarios will result in out of sample values of advertising. The exact magnitude of effects of out of sample advertising are driven by the functional form assumption in the model.¹³

6.3 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 co-operative's ideal. As in the demand estimation, the business stealing effects of advertising are swamped by the category expansive effects, cooperation should lead to an increase in total category advertising.

For illustrative purposes Figure 13 shows the computed benchmark versus co-operative advertising choice in the Atlanta DMA in the co-operative scenarios described above, with the left panel being the Eli Lilly/GSK co-operative and the right panel being the full category advertising co-operative. The Eli Lilly/GSK co-operative advertises about

¹³Alternative specifications with quadratic and square root functions of advertising were tried and produced qualitatively and quantitatively very similar results.

50% more in total than the sum of Prozac and Paxil in the competitive equilibrium. The full category co-operative advertises almost four times as much as the competitive equilibrium.

6.4 Quantities and Profits

Figure 14 illustrates the difference in profits between the benchmark and the co-operative situations. In the left panel, we see that in the Eli Lilly/GSK co-operative, the included products see roughly 10% higher profits than they do combined in the benchmark. In the right panel, we see that the full category profits increase by about 14% by the end of the period. Figure 15 compares the category share of the population in the Atlanta DMA between the co-operatives and the benchmark. In the Eli Lilly/GSK co-operative the category is about 6% larger than it is under competition. In the full co-operative, the category is almost 18% larger by the end of the period than it is under competition.

6.5 Discussion

While market expansion and increasing profits in the counterfactual would be viewed as welfare increasing in many consumer goods markets, there are a few reasons for us to take caution in drawing conclusions about social welfare in the context of antidepressants. First, many prescriptions are covered by insurance. While many people are getting prescribed and incurring minimal if any cost, the insurance system pays out a significant price. It is possible that the new prescriptions are not justified by the societal cost. However, as many physicians see depression as an under treated condition, the total welfare benefits could also be very high. Second, if I have missed important price or detailing complementarities, it might be the case that all increased profits are competed away after the co-operative sets higher advertising. If this is the case, the welfare effect is also ambiguous. The balance of societal costs and benefits, while very interesting and important, is not identified in this study and is certainly worthy of further research.

7 Conclusions

Using data from the antidepressant market, I find that television advertising has significant positive spillovers. These effects are identified using the discontinuity in advertising generated by the borders of television markets. The strategy proves important, as failing to consider endogenous firm choices of advertising leads to over-statements 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 where there are positive shocks to rival advertising. To size 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 co-operative 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% and their own profits by 10%. Meanwhile, a full industry co-operative would set advertising four times as high as is observed in equilibrium and would increase industry shares by 18% and profits by 14%.

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. While there is a brand effect, it is short lived while the category expansion effect is more persistent. Especially for conditions that are seen as under treated, 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 might be allowed by regulators. In the absence cooperation, it is important for firms to 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, it is important for marketers and econometricians to consider the possibility of positive spillovers when building models of advertising impacts on supply and demand.

References

- [1] Aitken, M.L., Berndt, E.R., Bosworth, B., Cockburn, I.M., Frank, R., Kleinrock, M., & Shapiro, B.T. (2013). The regulation of prescription drug competition and market responses: patterns in prices and sales following loss of exclusivity. NBER Working Paper #19487.
- [2] Anderson, E. T., & Simester, D. (2013). Advertising in a competitive market: the role of product standards, customer learning, and switching costs. *Journal of Marketing Research*, 50(4), 489-504.
- [3] Anderson, I., (2000). Selective serotonin reuptake inhibitors versus tricyclic antidepressants: a meta-analysis of efficacy and tolerability. *Journal of Affective Disorders*, 58, 19-36.(1)
- [4] Avery, R. J., Eisenberg, M., & Simon, K. I. (2012). The impact of direct-to-consumer television and magazine advertising on antidepressant use. *Journal of Health Economics*, 31(5), 705-718.
- [5] Bagwell, K. (2007). *The economic analysis of advertising*. Handbook of industrial organization, 3, 1701-1844.
- [6] Bayer, P., Ferreira, F., & McMillan, R. (2007). A unified framework for measuring preferences for schools and neighborhoods. *Journal of Political Economy*, 115(4), 588-638.
- [7] Berndt, E. R. (1991). *The practice of econometrics: classic and contemporary*. Reading, MA: Addison-Wesley.
- [8] Berndt, E. R., Bui, L., Reiley, D. R., & Urban, G. L. (1995). Information, marketing, and pricing in the US antiulcer drug market. *The American Economic Review*, 85(2), 100-105.
- [9] Berndt, E. R., Bui, L., Reiley, D. R., & Urban, G. L. (1997). The roles of price, quality, and marketing in the growth and composition of the antiulcer drug market. *The economics of new goods* (NBER Studies in Income and Wealth, vol. 58), Bresnahan, T. F., & Gordon, R. J. (Eds.). (1997), (Vol. 58). University of Chicago Press., pp. 277-322.
- [10] Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, 25(2), 242-262.
- [11] Black, S. (1999). Do better schools matter? Parental valuation of elementary education. *Quarterly Journal of Economics*, 114(2), 577-599.

- [12] Busch, S., Golberstein, E., & Meara, E. (2011). The FDA and ABCs: The unintended consequences of antidepressant warnings on human capital. NBER Working Paper # 17426.
- [13] Card, D., & Krueger, A. (1994). Minimum wages and employment: a case study of the New Jersey and Pennsylvania fast food industries. *American Economic Review*, 84(4), 772-793.
- [14] Ching, A. (2010). A dynamic oligopoly structural model of the prescription drug market after patent expiration. *International Economic Review*, 51(4), 1175-1207.
- [15] Ching, A., Erdem, T., & Keane, M.P. (2009). The price consideration model of brand choice. *Journal of Applied Econometrics*, 24(3), 393-420.
- [16] Ching, A., & Ishihara, M. (2012). Measuring the informative and persuasive roles of detailing on prescribing decisions. *Management Science* 58(7), 1374-1387.
- [17] Dickstein, M. (2014). Efficient provision of experience goods: evidence from antidepressant choice. Stanford University working paper.
- [18] Donohue, J. M., & Berndt, E. R. (2004). Effects of direct-to-consumer advertising on medication choice: The case of antidepressants. *Journal of Public Policy & Marketing*, 23(2), 115-127.
- [19] Donohue, J. M., Berndt, E. R., Rosenthal, M., Epstein, A. M., & Frank, R. G. (2004). Effects of pharmaceutical promotion on adherence to the treatment guidelines for depression. *Medical Care*, 42(12), 1176-1185.
- [20] Dorfman, R., & Steiner, P. O. (1954). Optimal advertising and optimal quality. *The American Economic Review*, 44(5), 826-836.
- [21] Dube, Arindrajit, T. William Lester, and Michael Reich. "Minimum wage effects across state borders: Estimates using contiguous counties." *The Review of Economics and Statistics* 92.4 (2010): 945-964.
- [22] Dubé, J. P., & Hitsch, G. J., & Manchanda, P. (2005). An empirical model of advertising dynamics. *Quantitative Marketing and Economics*, 3(2), 107-144.
- [23] Duggan, M., & Morton, F. M. S. (2006). The distortionary effects of government procurement: evidence from Medicaid prescription drug purchasing. *The Quarterly Journal of Economics*, 121(1), 1-30.
- [24] Ellison, G., & Ellison, S. F. (2011). Strategic entry deterrence and the behavior of pharmaceutical incumbents prior to patent expiration. *American Economic Journal: Microeconomics*, 3(1), 1-36.
- [25] Ellison, S. F., & Wolfram, C. (2006). Coordinating on lower prices: pharmaceutical pricing under political pressure. *The RAND Journal of Economics*, 37(2), 324-340.
- [26] Filson, D. (2012). A markov-perfect equilibrium model of the impacts of price controls on the performance of the pharmaceutical industry. *The RAND Journal of Economics*, 43(1), 110-138.
- [27] Goeree, M. S. (2008). Limited information and advertising in the US personal computer industry. *Econometrica*, 76(5), 1017-1074.
- [28] Holmes, Thomas J. "The effect of state policies on the location of manufacturing: evidence from state borders." *Journal of Political Economy* 106.4 (1998): 667-705.

- [29] Iizuka, T. (2004). What explains the use of direct-to-consumer advertising of prescription drugs?. *The Journal of Industrial Economics*, 52(3), 349-379.
- [30] Iizuka, T., & Jin, G. Z. (2005). The effect of prescription drug advertising on doctor visits. *Journal of Economics & Management Strategy*, 14(3), 701-727.
- [31] Iizuka, T., Jin, G. Z. (2005). Direct to consumer advertising and prescription choice. Working Paper Available at SSRN 700921.
- [32] Ito, K. (2012). Do consumers respond to marginal or average price? Evidence from nonlinear electricity pricing. *American Economic Review*, 104(2), 537-563.
- [33] Jayawardhana, J. (2013). Direct-to-consumer advertising and consumer welfare. *International Journal of Industrial Organization*, 31(2), 164-180.
- [34] Kravitz, R. L., Epstein, R. M., Feldman, M. D., Franz, C. E., Azari, R., Wilkes, M. S., Hinton, L., & Franks, P. (2005). Influence of patients' requests for direct-to-consumer advertised antidepressants. *JAMA: the journal of the American Medical Association*, 293(16), 1995-2002.
- [35] Lewis, R., Nguyen, D., (2012). Wasn't that ad for an iPad? display advertising's impact on advertiser- and competitor-branded search. Yahoo Research Working Paper.
- [36] Lewis, R. A., & Rao, J. M. (2012). On the near impossibility of measuring advertising effectiveness. Yahoo Research Working paper.
- [37] Liu, Q. & Gupta, S. (2011). The impact of direct-to-consumer advertising of prescription drugs on physician visits and drug requests: empirical findings and public policy implications. *International Journal of research in Marketing*, 28, 205-217.
- [38] Liu, Q., Gupta, S., Venkataraman, S., & Liu, H. (2013). An empirical model of drug detailing: dynamic competition and policy implications. Purdue University working paper.
- [39] Manchanda, P., Rossi, P., & Chintagunta, P. (2004). Response modeling with non-random marketing mix variables. *Journal of Marketing Research*, 41(4), 467-478.
- [40] Murphy, M. J., R. L. Cowan, and L. I. Sederer (2009): *Blueprints: Psychiatry*, Philadelphia, PA: Lippincott Williams Wilkins, 5 ed.
- [41] Narayanan, S., Desiraju, R., & Chintagunta, P. K. (2004). Return on investment implications for pharmaceutical promotional expenditures: The role of marketing-mix interactions. *Journal of Marketing*, 68(4), 90-105.
- [42] Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, 1417-1426.
- [43] Nerlove, M., & Arrow, K. J. (1962). Optimal advertising policy under dynamic conditions. *Economica*, 29(114), 129-142.
- [44] Rosenthal, M. B., Berndt, E. R., Donohue, J. M., Epstein, A. M., & Frank, R. G. (2003). Demand effects of recent changes in prescription drug promotion. Henry J. Kaiser Family Foundation.
- [45] Sahni, N. (2013). Advertising spillovers: field experimental evidence and implications for the advertising sales-response curve. Stanford University working paper.

- [46] Stremersch, S., Landsman, V., & Venkataraman, S. (2012). The relationship between DTCA, drug requests, and prescriptions: uncovering variation in specialty and space. *Marketing Science*, 32(1), 89-110.
- [47] Wittink, D. R. (2002). Analysis of ROI for pharmaceutical promotion (ARPP). *Association of Medical Publications*, (September 18, 2002), [available at <http://www.vioworks.com/clients/amp>].
- [48] Wosinska, M. (2002). Just what the patient ordered? Direct-to-consumer advertising and the demand for pharmaceutical products. Direct-to-Consumer advertising and the demand for pharmaceutical products (October 2002). HBS Marketing Research Paper, (02-04).
- [49] Wosinska, M. (2005). Direct-to-consumer advertising and drug therapy compliance. *Journal of Marketing Research*, 42(3), 323-332.

Tables and Figures

Figure 1: Antidepressant Commercials Relative to FDA Memo

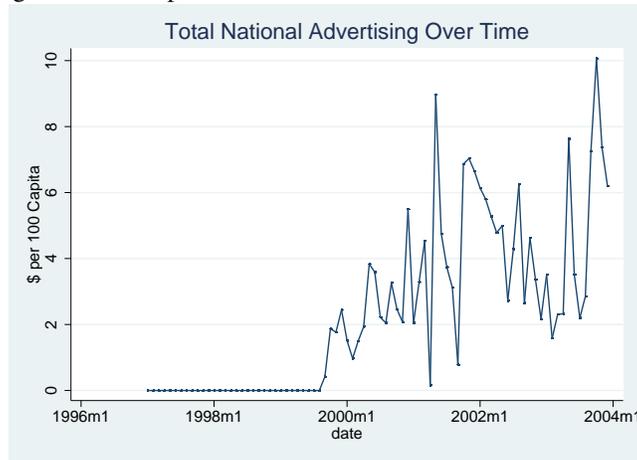


Figure 2: Antidepressant Revenues 1996-2008

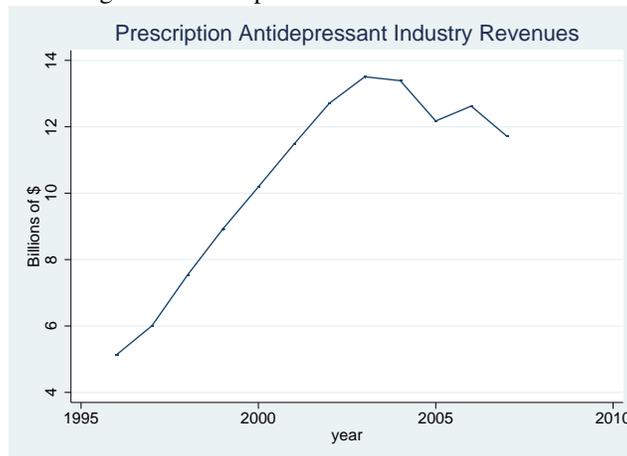


Table 1: Descriptive Statistics: Advertising

	Mean	Q25	Median	Q75
DTC per 100 capita	0.782	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

	Mean	Q25	Median	Q75
DMA's	101			
DMA Population	2340774	903090	1469823	2622567

Figure 3: Variation Across Three Markets in Advertising

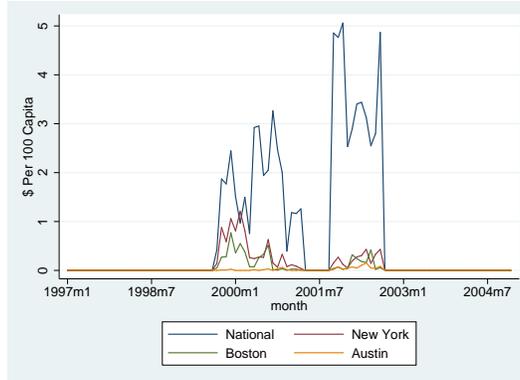


Figure 4: Full Sample: Top 101 DMAs

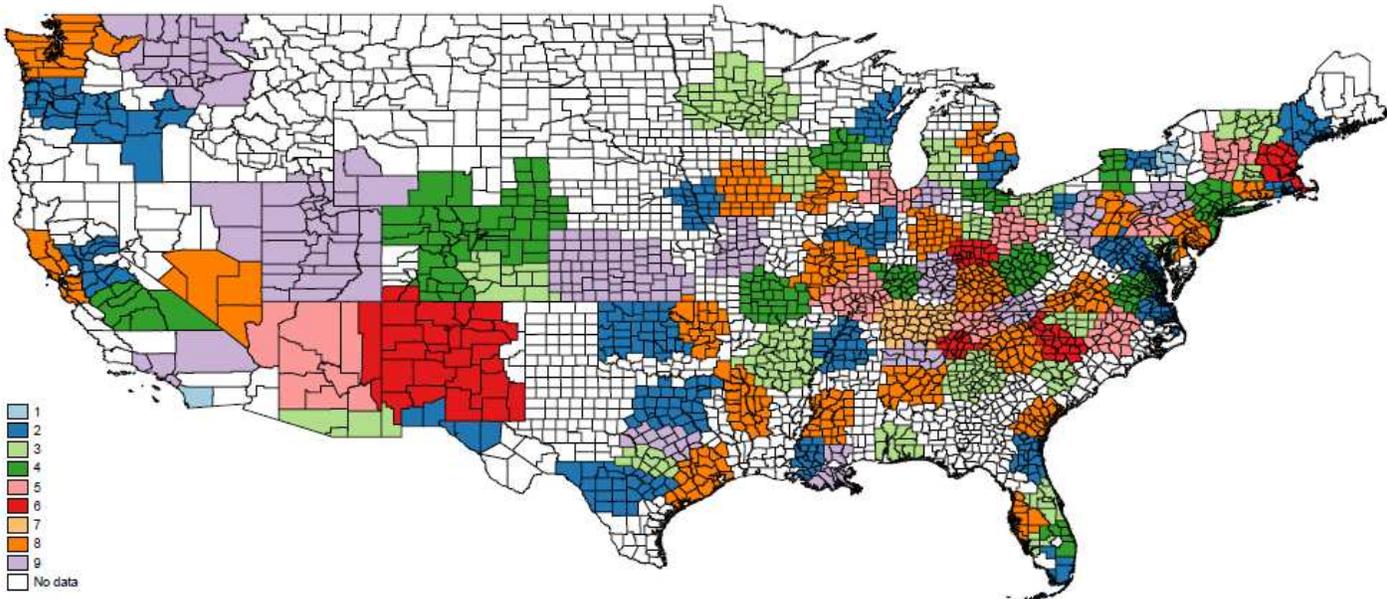


Figure 5: Ohio and Its DMAs

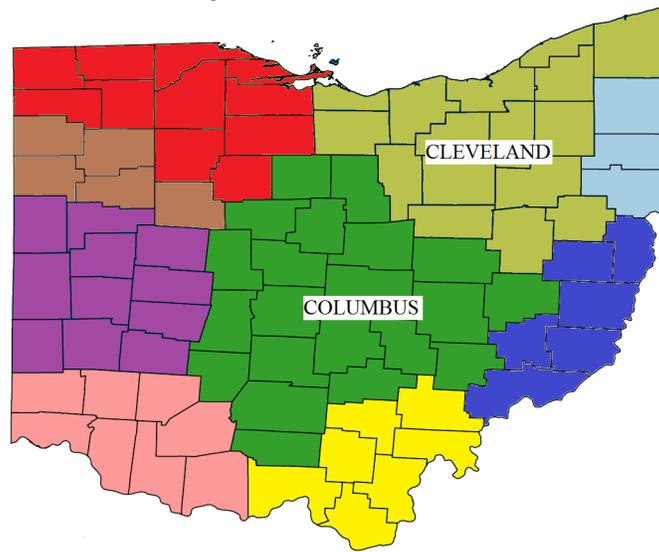


Figure 6: Border Sample: Counties on the Borders of the Top 101 DMAs

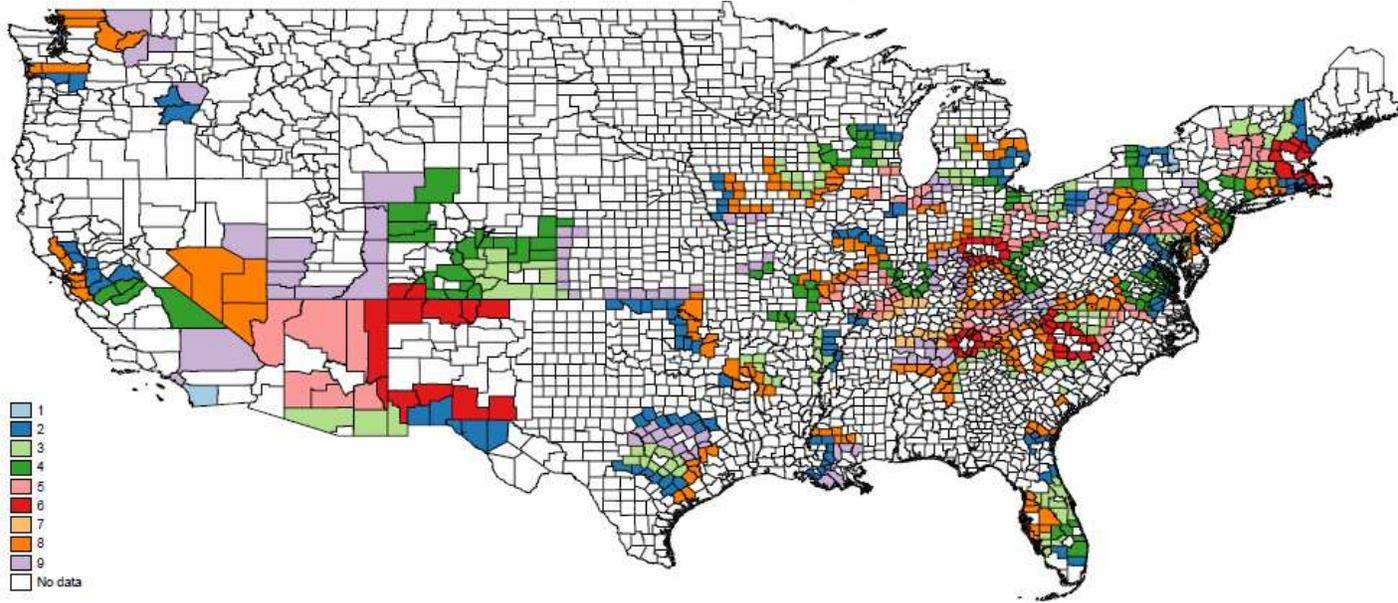


Figure 7: Variation in Log DTC Net of Fixed Effects, 14% Zeros

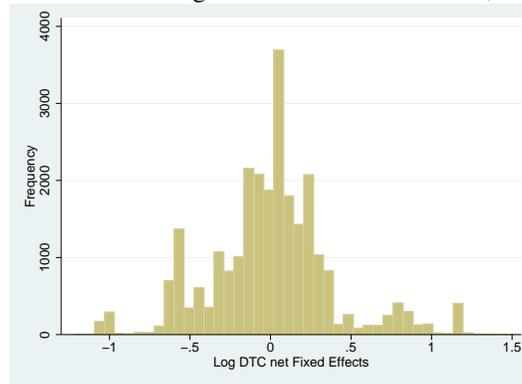


Table 2: The Effect of Own and Rival Advertisements on Sales

VARIABLES	(1) $\log(Q)$
lagged $\log(Q)$	0.334*** (0.00746)
DTC	0.0240*** (0.00621)
DTC^2	-0.00216* (0.00113)
DTC_{rival}	0.0164*** (0.00266)
DTC_{rival}^2	-0.000938*** (0.000252)
$DTC \times DTC_{rival}$	-0.00134** (0.000631)
Product-Border-Time	yes
Product-Border-DMA	yes
Observations	316,428
R-squared	0.955

*** p<0.01, ** p<0.05, * p<0.1

Product-DMA clustered standard errors in parentheses

Table 3: Descriptive Statistics: Border Sample, 1997-2003

	Mean	Q10	Median	Q90
Number of Border Experiments	153			
Number of DMAs	97			
$LOGDTC_{product}$	0.190	0	0	1.033
$LOGDTC_{nest}$	0.447	0	0	1.682
$LOGDTC_{market}$	0.817	0	0.921	1.987

$LOGDTC$: log of one plus dtc expenditures per 100 capita

All are defined at the experiment-DMA-month level

Table 4: Demographic Descriptive Statistics: Border Sample, 1997-2003

Variable	Mean	SD	Min	Max
<i>PctBlack</i>	0.0744	0.0994	0	0.628
<i>PctHispanic</i>	0.0647	0.102	0.00210	0.774
<i>PctAsian</i>	0.0137	0.0245	0.00034	0.290
<i>PctUrban</i>	0.00086	0.00083	0	0.00715
<i>PctUninsured</i>	0.00043	0.00078	0	0.0104
<i>PctOver45</i>	0.378	0.0501	0.217	0.571
<i>PctMale</i>	0.492	0.0101	0.466	0.569
<i>PctEmployment</i>	0.456	0.0560	0.213	0.619
<i>Income</i>	\$23,992	\$5,691	\$11,044	\$55,157

Demographic information is not available on a monthly basis. *PctUrban* and *PctUninsured* are defined at the experiment-DMA level only, using data from 2000, as these variables are only available every ten years with the census. All other demographic variables are defined at the experiment-DMA-year level.

Table 5: Results of Base Model

VARIABLES	Category Level	Subcategory Level	Product Level
<i>adstock</i>	0.0494*** (0.00769)	0.00662 (0.00828)	0.0250** (0.00919)
<i>xPctBlack</i>	-0.00300 (0.00920)	0.00617 (0.0126)	-0.000640 (0.00984)
<i>xPctHispanic</i>	-0.0112 (0.0138)	-0.0152 (0.0132)	0.0111 (0.0112)
<i>xPctAsian</i>	0.00023 (0.0122)	-0.00793 (0.0180)	-0.0126 (0.0131)
<i>xPctUrban</i>	0.00421 (0.0109)	0.00726 (0.0155)	0.00818 (0.0202)
<i>xPctUninsured</i>	-0.0150 (0.0155)	0.00140 (0.0210)	-0.0302 (0.0215)
<i>xPctOver45</i>	0.0153* (0.00713)	0.0182 (0.0110)	-0.00511 (0.0105)
<i>xPctMale</i>	-0.00366 (0.00731)	0.00869 (0.00719)	-0.0175* (0.00710)
<i>xEmployment</i>	0.00336 (0.00797)	-0.0150 (0.0112)	-0.0035 (0.0104)
<i>xIncome</i>	-0.0134 (0.00951)	0.0210 (0.0148)	-0.0184 (0.0118)
<i>persistence, λ</i>	0.680*** (0.0309)	0.279*** (0.0120)	0.324*** (0.0139)
Observations	22,592	92,238	139,305
R-squared	0.952	0.936	0.961

* p<0.05, ** p<0.01, *** p<0.001

Level-DMA clustered standard errors in parentheses. Demographic variables are log normalized with mean zero and standard deviation of 1 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.

Figure 8: Impulse Response Effect of Zoloft Advertisement on Own and Total Prescriptions

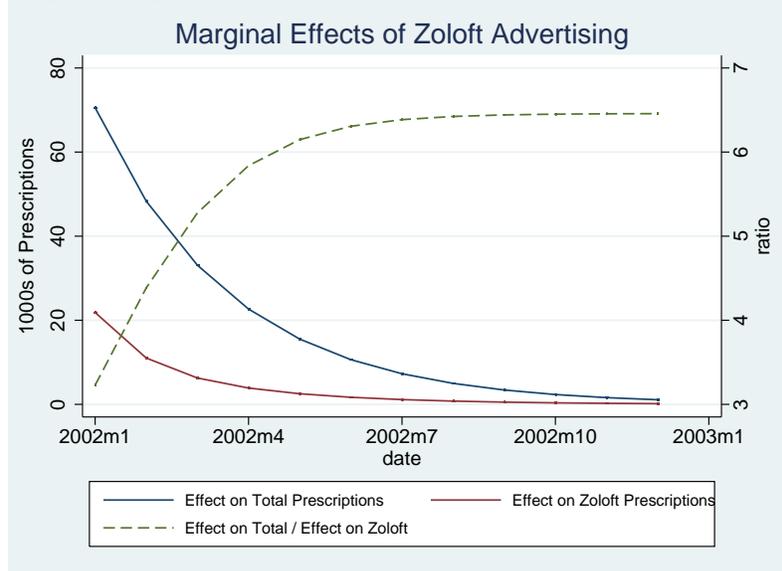


Table 6: Main Results: Importance of the Border Strategy

VARIABLES	No Border	No Border	Border
$adstock_{category}$	0.0118*** (0.00113)	0.0453*** (0.00234)	0.0490*** (0.00665)
$\lambda_{category}$	0.968*** (0.00523)	0.715*** (0.0202)	0.680*** (0.0309)
$adstock_{subcategory}$	0.0326*** (0.00221)	0.00243 (0.00324)	0.00877 (0.00795)
$\lambda_{subcategory}$	0.973*** (0.00122)	0.699*** (0.0138)	0.279*** (0.0120)
$adstock_{product}$	0.0652*** (0.00669)	0.0203*** (0.00524)	0.0254** (0.00897)
$\lambda_{product}$	0.961*** (0.00184)	0.594*** (0.0115)	0.324*** (0.0139)
Fixed Effects		X	X

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Level-DMA clustered standard errors in parentheses. Demographic variables are log normalized with mean zero and standard deviation of 1 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.

Table 7: Short Run Advertising Elasticities

Products	Paxil	Paxil CR	Prozac	Prozac Weekly	Wellbutrin SR	Wellbutrin XL	Zoloft	Outside Option
Paxil	0.037	0.019	0.021	0.019	0.020	-	0.021	-0.023
Paxil CR	0.016	0.029	0.016	0.016	0.012	0.010	0.016	-0.015
Prozac	0.0092	-	0.020	0.0080	0.0097	-	0.0092	-0.011
Prozac Weekly	0.0088	-	0.0088	0.018	0.0068	-	0.0088	-0.0080
Wellbutrin SR	0.014	-	0.014	0.012	0.021	-	0.014	-0.014
Wellbutrin XL	0.017	0.017	0.017	0.017	0.019	0.035	0.017	-0.018
Zoloft	0.013	0.013	0.013	0.013	0.013	0.010	0.027	-0.015

Figure 9: Marginal Revenue Curves Under Various Scenarios

Marginal Revenue of Advertising for Zoloft

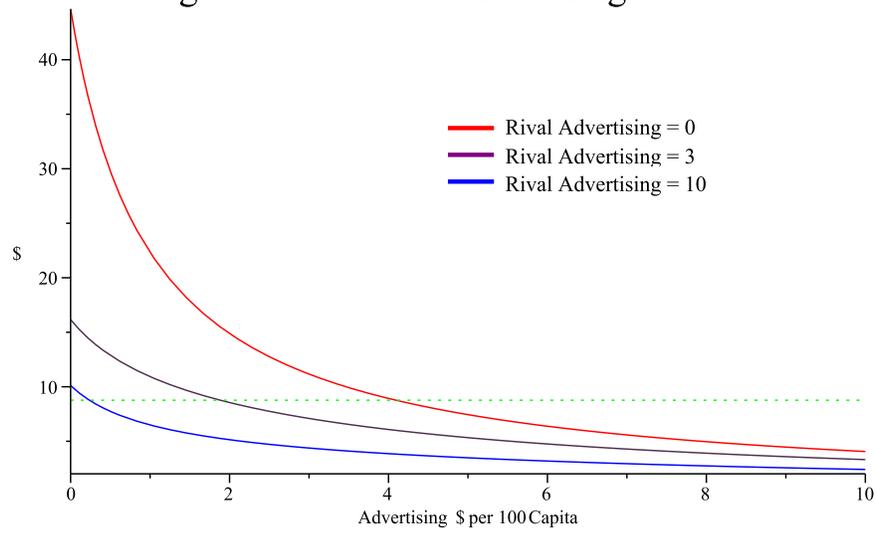


Table 8: Evidence of Free Riding By Firms

VARIABLES	<i>DTC</i>	<i>Log(DTC)</i>	$\mathbf{1}(DTC>0)$	$\mathbf{1}(DTC>0)$
Rival <i>DTC</i>	-0.0296*** (0.00240)			-0.00334* (0.00160)
<i>Log(RivalDTC)</i>		-0.0290*** (0.00691)		
$\mathbf{1}(RivalDTC>0)$			-0.0685*** (0.0162)	
Observations	39,414	39,414	39,414	39,414

*** p<0.01, ** p<0.05, * p<0.1

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

Figure 10: MPE Simulation Advertising - Atlanta DMA

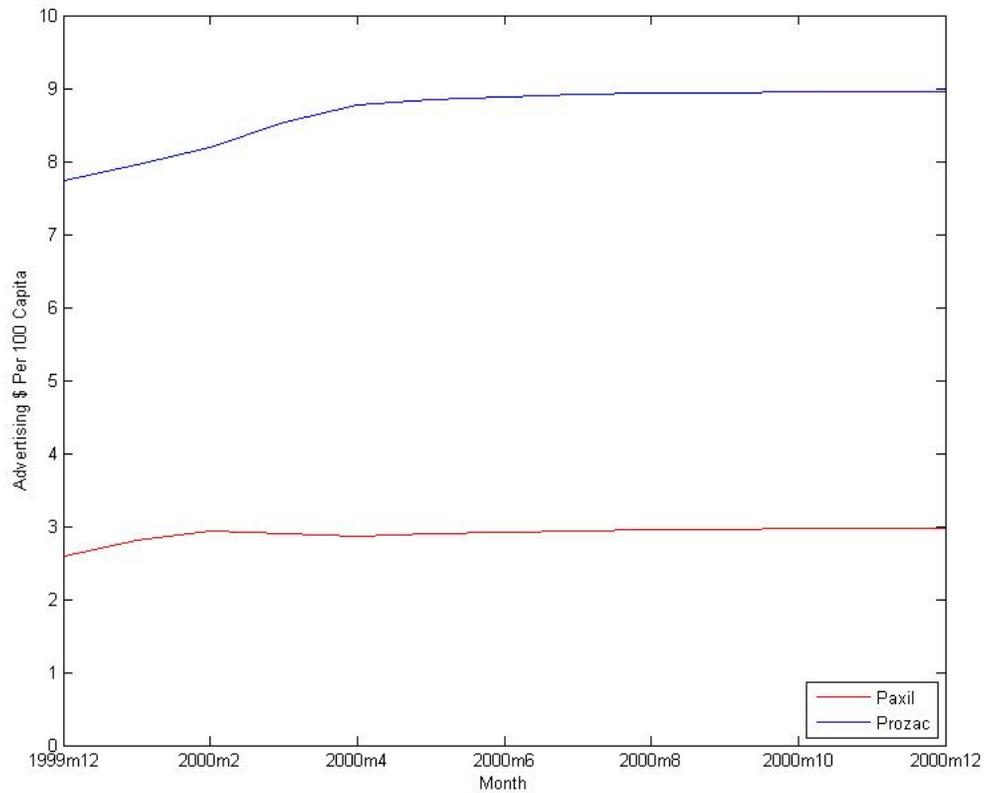


Figure 11: Realized Advertising - Atlanta DMA

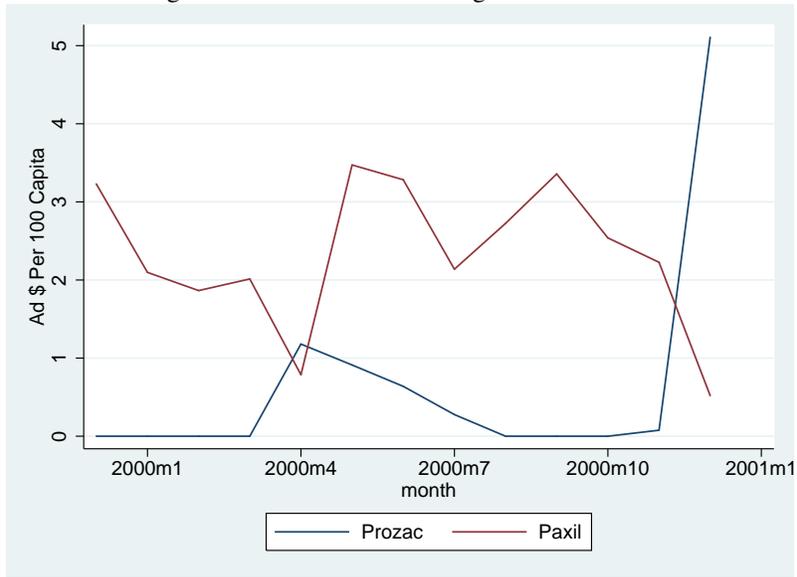


Figure 12: MPE Simulation Profits - Atlanta DMA

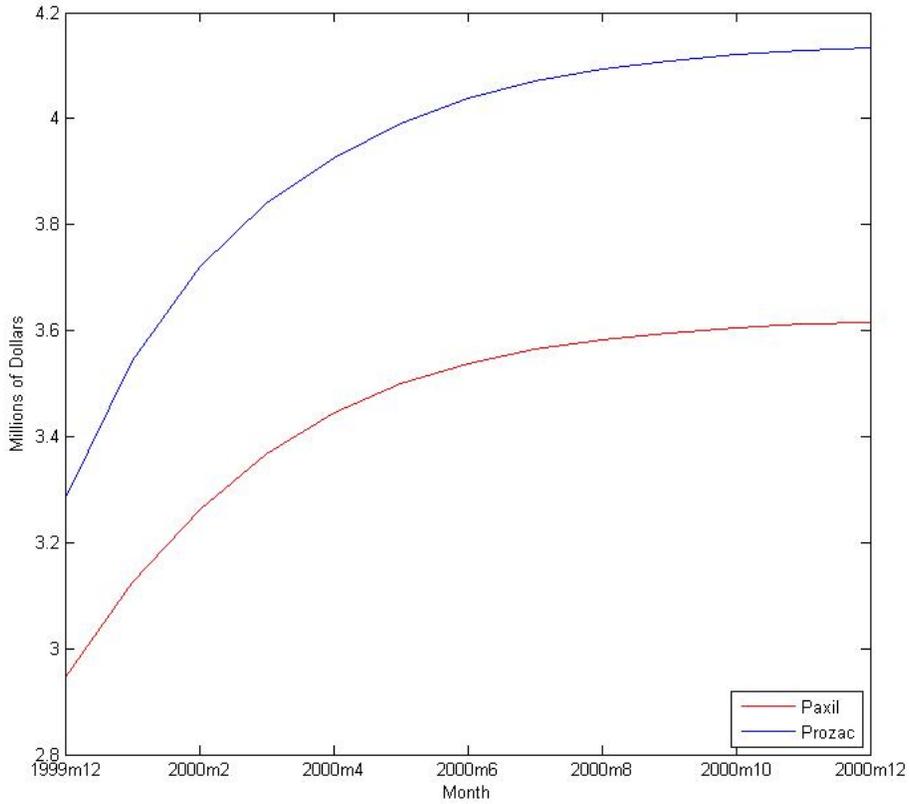


Figure 13: Paxil/Prozac and Full Co-operative versus Competitive Total Advertising - Atlanta DMA

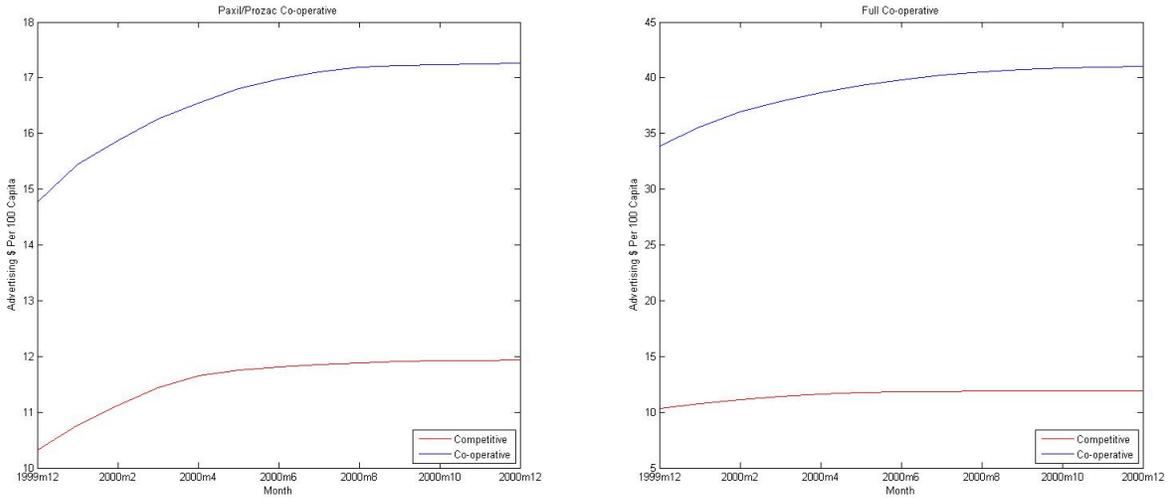


Figure 14: Paxil/Prozac Co-operative versus Competitive Total Profits - Atlanta DMA

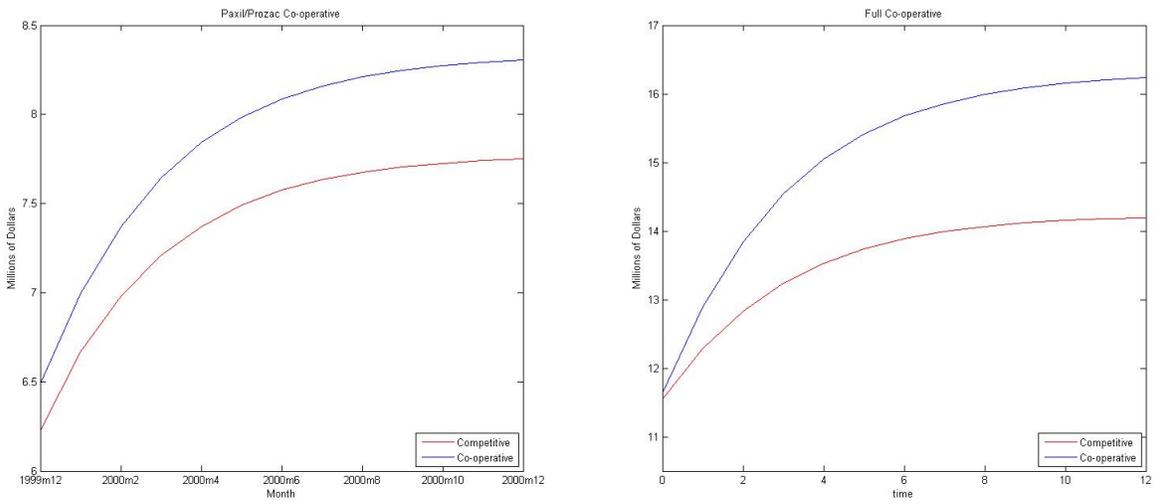
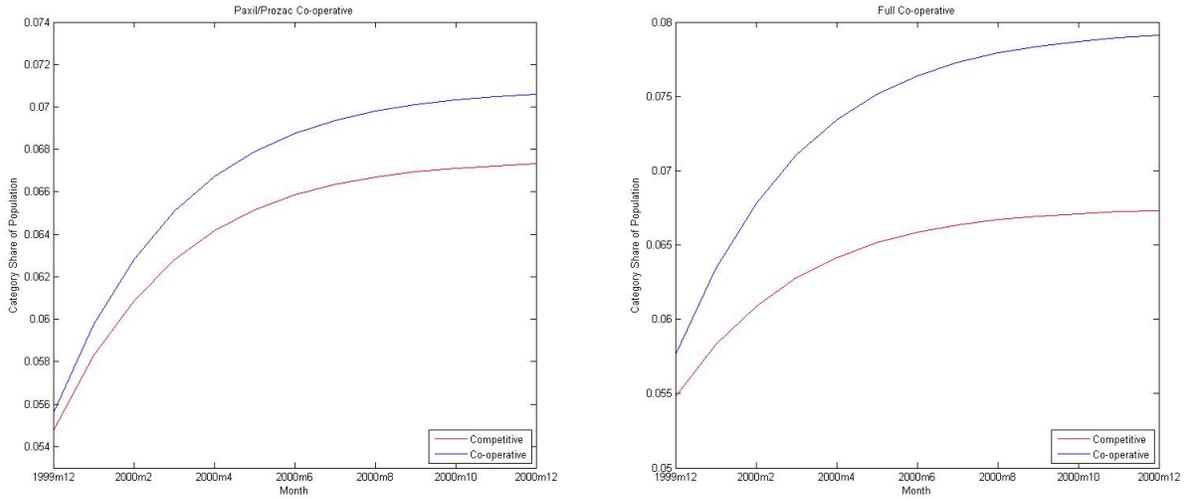


Figure 15: Paxil/Prozac Co-operative versus Competitive Category Share - Atlanta DMA



Appendix A - Orthogonality of Pricing and Detailing Decisions

A.1 Pricing

The regression of television commercials on prices is shown in Table 9. As can be seen the point estimate is very small and insignificant with respect to both own and cross advertising. The average unit price of a branded drug is about \$3.24 over the course of the sample and the average DTC per capita for advertising products after September 1999 is \$0.65. The point estimate suggests that raising DTC per capita by \$1 is associated with a price decrease of \$0.05. This is very small economically. Interestingly, a time trend, a product fixed effect and a dummy for patent expiration can explain prices with R squared larger than 0.99. Prices seem quite sticky, especially relative to advertising in this market.

Table 9: Predicting Prices with Advertising

VARIABLES	(1) realunitprice
<i>DTC</i>	-0.0518 (0.0437)
<i>RivalDTC</i>	-0.0332 (0.0248)
time	0.00963*** (0.00313)
expired	-0.210 (0.149)
Product FEs	X
Observations	1188
R-squared	0.994
*** p<0.01, ** p<0.05, * p<0.1	
Product Clustered Standard errors in parentheses	

A.2 Detailing

To address the potential omitted variables bias problem associated with physician detailing, I use two data sources to show that physician detailing is uncorrelated with DTC advertising. First, I use national level detailing information from IMS health and exploit the law change that made DTC advertising feasible for pharmaceutical firms starting in 1999. Using this data, I show that there is no trend break in firm detailing associated with the policy change, while there is an enormous trend break in DTC, from zero to highly positive. Second, I employ data from ImpactRx that follows a panel of physicians through time, measuring the amount of detailing time was spent on them by drug. The data includes a panel of 2134 general practitioners and 167 psychiatrists that can be matched with the DTC and prescribing data. These physicians are not representative of the population of physicians, as they are drawn disproportionately from the 40th percentile and higher in the prescribing distribution. ImpactRx does this because those below the 40th percentile are much less likely ever to be detailed. While the 2301 physicians make up less than 1% of the population of physicians, they are likely make up significantly more of the prescription share of the physician population. For general practitioners, the panel begins in 2001 and for psychiatrists, the panel begins in 2002 and includes geographic information. As such, the data is matched with a portion of the data used in the main analysis to test the robustness of the identification to the inclusion of detailing. Detailing is measured in number of detail visits per product per physician per month. With 101 DMAs in the data, this makes each DMA have an average of about 23 physicians in this sample.

Using this data, I use two approaches to address the omitted variables bias concern. First, I aggregate the local detailing measure to the DMA level and show that DTC and local detailing are not correlated conditional on the fixed effects included in the model. Similarly, I aggregate the data to the border experiment level and show that this lack of correlation extends to the border regions. Next, I include the border-experiment level detailing data in the model for the limited months and experiments that can be matched and show that the exclusion of detailing has no effect on the estimates of DTC effectiveness.

Table 10 presents the results of the national-level analysis using IMS national detailing visit numbers. The regressions are specified using product specific time trends, product fixed effects, a dummy to indicate that the product is in the market, a dummy to indicate that the product’s patent has expired and a dummy for “post” September 1999, when the final FDA guidance was released. The coefficient on “post” will identify the average break in trend of detailing at the moment the law changes. The coefficient is small and positive, but insignificant. In contrast, using the same specification but with DTC as the dependent variable in the second column, the “post” dummy shows a very large positive and significant effect, as firms move from zero DTC advertising at the national level to a significant amount of DTC advertising. Even at the aggregate national level, it appears as though this major law change affecting firm strategy has very little effect on the path of firm detailing decisions.

Table 10: Did the DTC Law Change Result in a Detailing Trend Break?

VARIABLES	National Details	National DTC
<i>PostSept99</i>	0.672 (0.796)	472*** (124)
Product Time Trend	X	X
Observations	992	992
R-squared	0.932	0.266
*** p<0.01, ** p<0.05, * p<0.1		

Standard errors in parentheses. Included in each specification are product fixed effects, product specific time trends, a dummy indicating a product’s patent has expired and a dummy indicating that the product is in the market. Average detailing per month in the sample is 16.807. Average national DTC per month in the sample is 427.750

The next strategy employed to assess the potential for bias is to aggregate the physician-level detailing from ImpactRx to the DMA level. Since DTC advertising is decided at the DMA level and no more local, I want to test if detailing decisions are sensitive to DTC decisions at the DMA level. I regress own local DTC and rival local DTC, measured in dollars per hundred capita on product detailing visits in the DMA. Similar to the main specification, I include product-quarter and product-DMA fixed effects. I want to know within time period and within DMA, if higher DTC for one product associated with higher detailing. The results of the regression are presented in the first column of Table 11. As it turns out, there is no statistically significant relationship between either own or rival DTC and detailing decisions. This is consistent with conversations with managers who said that detailing and DTC decisions are made in different divisions of the firm and coordination very difficult. While the estimates are not statistically significant, it is worth noting that even if they were, the magnitudes of the point estimates are very small. The standard deviation of local DTC campaigns is 0.17. Taken at face value, an increase in own local DTC of one standard deviation is associated with an increase in detailing visits for doctors in this sample in the DMA by about 0.1. Extrapolating this point based on the size of the sample, the point estimate would imply an average of about 1 additional detailing visit per 230 physicians associated with a one standard deviation increase in local DTC. The point estimate on rival advertising is even smaller. This provides evidence that widespread strategies to undercut rival DTC with detailing campaigns are not prevalent.

In the second column of Table 11, the same exercise is run at the border-experiment level. The results are broadly similar, though these results should be taken with a grain of salt as each observation is based on far fewer physicians at this narrow level of analysis. However, there is no evidence that detailing efforts change sharply at borders where one side has significantly higher DTC.

Table 11: Are Detailing Campaigns Coordinated with DTC?

VARIABLES	DMA Level Detailing Visits	Border-Experiment Level Detailing Visits
Local DTC	0.662 (0.475)	0.579 (0.490)
Rival Local DTC	0.184 (0.177)	0.0326 (0.108)
Observations	22,589	18,622
R-squared	0.764	0.789

*** p<0.01, ** p<0.05, * p<0.1

Product-DMA clustered standard errors in parentheses. Column 1 presents DMA level analysis while Column 2 provides border experiment level analysis. They contain the same fixed effects as those in the demand models of the same level.

The final strategy used to think about this problem is to directly incorporate local detailing data into the main analysis. Detailing data is aggregated to the product-border-DMA-time level as with the main specification. Since the physician panel in the detailing data is smaller in both number of physicians and in length of time, the matched sample will be substantially smaller and over a shorter time frame than the main specification. As such, there is no longer sufficiently large T to ensure that the lagged dependent variable Nickel bias is small. The persistence parameters will all be biased downward in these specifications. The main goal of this exercise is to see if the coefficients on DTC are robust to the inclusion of detailing.

The effect of including detailing in this model is demonstrated in Tables 12 and 13 . Both tables include all matched data from 2001 onwards. The regressions in Table 12 include the log of detailing visits while the regressions in Table 13 do not. Table 12 presents specifications that both include and exclude interaction terms between DTC and detailing. As we can see, the effects of DTC are not significantly affected by the exclusion of detailing or the interactions of DTC with detailing. As expected, the persistence parameters are all much smaller than in the main model in Table 5 due to the smaller number of time periods.

These exercises show that over the course of the time period where detailing data is available in this class, detailing is not significantly correlated with DTC at either the DMA level or the border experiment level. Further, including detailing in the main model does not change the main parameters of interest. While this paper makes no claims of identifying the causal effects of detailing, it seems as though it poses a minimal threat to the identification of the causal effect of DTC using the border strategy.

Appendix B - Placebo Test

With a difference-in-differences model, the assumption of parallel trends in the outcome variable absent the treatment is required for a valid estimation. One way to assess the validity of this assumption is through the use of a placebo test. In this case, I will use advertising for over-the-counter (OTC) sleep aids as a placebo treatment. This is an ideal placebo for two reasons: first, it varies at the same level as antidepressant advertising- at the DMA month. Next, OTC sleep aids need not be prescribed by a physician, so we should not expect a “going to the doctor” effect of advertising

Table 12: Results of Limited Model Including Geographic Detailing from 2001 On

VARIABLES	Category Level	Category Level	Subcategory Level	Subcategory Level	Subcategory Level	Product Level	Product Level
<i>adstock</i>	0.0484*** (0.00638)	0.0474** (0.00730)	0.00345 (0.00922)	0.00710 (0.00968)	0.0300** (0.0103)	0.0314** (0.0105)	
<i>LogVisits</i>	0.00570 (0.00578)	0.00806 (0.00806)	-0.00471 (0.00568)	0.00029 (0.00733)	0.00236 (0.00358)	0.00378 (0.00362)	
<i>LogVisits</i> × <i>adstock</i>		0.00157 (0.00357)		-0.00682 (0.00485)		-0.00558 (0.00659)	
<i>persistence</i>	0.489*** (0.0497)	0.489*** (0.0496)	0.124*** (0.0146)	0.124*** (0.0146)	0.260*** (0.0153)	0.260*** (0.0153)	
Observations	9,861	9,861	42,860	42,860	77,891	77,891	
R-squared	0.960	0.960	0.940	0.940	0.968	0.968	

*** p<0.01, ** p<0.05, * p<0.1

Level-DMA clustered standard errors in parentheses. Demographic interactions included as in the main analysis but suppressed for expositional clarity.

Table 13: Results of Limited Model Without Geographic Detailing

VARIABLES	Category Level	Subcategory Level	Product Level
<i>adstock</i>	0.0491*** (0.00636)	0.00301 (0.00911)	0.0301*** (0.00103)
<i>persistence</i>	0.489*** (0.0497)	0.124*** (0.0146)	0.260*** (0.0153)
Observations	9,861	42,860	77,891
R-squared	0.960	0.940	0.968

*** p<0.01, ** p<0.05, * p<0.1

DMA clustered standard errors in parentheses. Demographic interactions included as in the main analysis but suppressed for expositional clarity.

to be present in OTC advertising. I will use the same identification strategy, but I will use OTC sleep aid advertising as a treatment. The results are presented in Table 14. None of the coefficients on OTC sleep aid advertising is statistically significant at the 5% level or economically important at any level.

Table 14: Results of Base Model with Placebo

VARIABLES	Category Level	Subcategory Level	Product Level
<i>OTCSLEEPAD</i>	0.0000565 (0.000397)	0.000471 (0.000482)	0.00218* (0.00113)
<i>persistence</i>	0.684*** (0.0304)	0.280*** (0.0116)	0.323*** (0.0139)
Observations	22,826	93,280	147,443
R-squared	0.948	0.935	0.960

*** p<0.01, ** p<0.05, * p<0.1

Level-DMA clustered standard errors in parentheses

Appendix C - Alternative Sample Selection

One might worry that the effect of advertising in the border sample counties differs systematically from the non-border counties or that the effect of advertising in rural areas would be much different than the effect of advertising in urban areas. To think about these concerns, I have repeated the analysis with several alternative sample selections.

C.1 The Urban Rural Divide

C.1.1 Without the Northeast Corridor and Other Urban Areas

It seems the less urban areas show very similar results to the full border sample. The category and product level effects are larger, but not significantly different from the full sample of borders.

Table 15: Results of Base Model without Northeast Corridor and Other Urban Areas

VARIABLES	Category Level	Subcategory Level	Product Level
<i>adstock</i>	0.0508*** (0.00825)	0.00874 (0.00904)	0.0279*** (0.00939)
<i>persistence</i>	0.689*** (0.0322)	0.255*** (0.0129)	0.313*** (0.0159)
Observations	17,400	67,976	99,101
R-squared	0.946	0.927	0.955
*** p<0.01, ** p<0.05, * p<0.1			
Level-DMA clustered standard errors in parentheses			

C.1.2 Only the Urban Border Counties

It seems the effects of advertising are a bit smaller in the more urban areas, except at the nest level. However, it seems more likely that the identifying assumption might fail in the more urban areas, as the borders are much closer to the central cities than in the more rural borders.

Table 16: Results of Base Model with only Northeast Corridor and Other Urban Areas

VARIABLES	Category Level	Subcategory Level	Product Level
<i>adstock</i>	0.0348*** (0.00951)	0.0191** (0.00940)	0.00707*** (0.0111)
<i>persistence</i>	0.648*** (0.0678)	0.387*** (0.0215)	0.381*** (0.0284)
Observations	5,775	25,550	41,586
R-squared	0.963	0.956	0.972
*** p<0.01, ** p<0.05, * p<0.1			
Level-DMA clustered standard errors in parentheses			

C.2 The “Anti” Border Sample

Here, I use all of the counties that are not in the border sample. For this sample, endogeneity should be the most severe. The results of this estimation should reveal the main drivers of the endogeneity problem that necessitates use of the border sample. If the main effects of advertising are much larger in this estimation, it would lead us to believe that advertising was strongly correlated with unobserved local demand shocks such as weather shocks, employment shocks or special seminars that are unobserved. These shocks, if positively correlated with advertising decisions would lead us to falsely conclude that the effect of advertising is high. If, on the other hand, the reverse causality concern- where firms target an advertising to sales ratio based on sales in the previous period, we would expect that in the anti-border sample the persistence parameter would be much higher. That is, lagged shares would not only be measuring the persistence of prescribing associated with advertising carry-over, they would also be controlling for the rule of thumb decision of firms to target DMAs that had large shares in the previous period. Results of the estimation are in Table

14. As can be seen, while the point estimates on advertising do not significantly change from the border sample, they are larger, and the persistence parameters are much (and significantly) larger. This suggests that the main source of endogeneity is reverse causality- if firms are following a rule of thumb advertising to sales ratio based on previous period sales, the lagged share variable will be biased upward as it controls for the firm’s rule of thumb rather than estimating persistence.

Table 17: Results of Base Model Using the “Anti”-Border Approach

VARIABLES	Category Level	Subcategory Level	Product Level
<i>adstock</i>	0.0508*** (0.00164)	0.0049 (0.00251)	0.0182*** (0.00439)
<i>persistence</i>	0.751*** (0.0259)	0.721*** (0.0128)	0.594*** (0.0115)
Observations	8,466	42,181	73,628
R-squared	0.981	0.984	0.984

*** p<0.01, ** p<0.05, * p<0.1

Level-DMA clustered standard errors in parentheses

C.3 Removing the State Borders

As described in the text, if tax rates cause a selection problem of physicians into DMAs, then results could be biased. To assess this possibility, DMA borders which coincide with state borders are removed from the sample. That removes roughly one third of the borders in the sample. The analysis is re-run using the set of DMA borders that do not coincide with state borders. Results are available in Table 17. The results are very similar and not statistically distinguishable from those in the full sample.

Table 18: Results of Base Model Without State Border Overlap

VARIABLES	Category Level	Subcategory Level	Product Level
<i>adstock</i>	0.0376*** (0.00764)	0.0131* (0.00783)	0.0256*** (0.00881)
<i>persistence</i>	0.683*** (0.0368)	0.271*** (0.0133)	0.327*** (0.0166)
Observations	15,975	65,097	97,720
R-squared	0.981	0.984	0.984

*** p<0.01, ** p<0.05, * p<0.1

Level-DMA clustered standard errors in parentheses

Appendix D - Primary Care Service Areas

As mentioned in section (3.1.1), measurement error could bias my estimates towards zero if patients are going to the doctor in different counties than where they watch television advertisements. The Dartmouth Center for Health Policy

Research has developed a Primary Care Service Area (PCSA) project which is the first national database of primary care resources for small areas. These areas were defined using Medicare claims data from 1999 and Census data from 2000. The service areas include a ZIP area with one or more primary care providers and any bordering ZIPs where the population largely gets their primary care from those physicians.

This database allows me to ask how many patients travel across DMA borders to seek their primary care. In particular, I can match this data to my prescribing data at the ZIP level. I can then see what percentage of each PCSA falls into a single DMA. Doing this I find that only about 1% of PCSAs cross DMA borders at all. Of those that do cross DMA borders, they do so only minimally. That is, the DMA holding the majority of a PCSA which crosses a border on average contains 97% of that PCSA. As such, measurement error bias should be minimal.

While this only holds for Medicare beneficiaries and primary care physicians, it should be noted that primary care physicians do make up a large portion, on the order of 80% of antidepressant prescriptions.¹⁴ For this to be a large measurement error issue, non-Medicare recipients must travel larger distances than Medicare recipients to their physicians. To the extent that this is true, it will lead to measurement error that biases all estimates of advertising effectiveness towards zero.

Appendix E - Selection of Households and Physicians to Counties

E.1

It might present a problem if households and physicians pre-disposed to be more affected by advertisements systematically select into border counties on the side of the border that receive more advertisements. This is a difficult level of selection to address. However, I do check for balance in characteristics across the borders. First, I check to see if physicians are more likely to bunch on the “high DTC” side of the border. To do this, I collected data from the Area Resource File from 2001 that contained information on the number of physicians in a particular county- I collected this data for both total number of physicians and non-federal physicians. To determine the “high DTC” side of the border, I totaled the local ads from 2001 for each side of a particular border. If one side saw more advertising in total in 2001, it is called the “high DTC” side. I then ran a t-test to see if the means of the numbers of physicians were the same on either side of the border. The t-test fails to reject that the number of physicians is the same on either side of the border, with $p=0.21$ for total doctors and $p=0.25$ for non-federal doctors.

Next, we might be concerned that the counties on the “high DTC” side of the border are systematically higher income or higher population than those on the “low DTC” side of the border. A t-test fails to reject that average income is the same across the border, with $p=0.47$. Similarly, a t-test for population fails to reject that the populations are the same across the border with $p=0.17$.

Results for all of these tests are presented in Table 18. It should be noted that the market fixed effects in the model allow for different levels in all of these variables so long as the trends are parallel and that the effectiveness of advertising is the same across the border. That there does not seem to be systematic differences in characteristics across the border further solidifies that this level of comparison is a reasonable one.

¹⁴<http://psychcentral.com/news/2011/08/08/more-antidepressants-prescribed-by-non-psychiatrists/28423.html>.

Table 19: Tests for Selection Across Borders

variable	high DTC	low DTC	t-stat	p-value
Physicians	855 (136)	715 (124)	1.25	0.215
Non-Fed Physicians	98.8 (12.9)	88.0 (11.3)	1.15	0.252
Income	24,162 (460)	23,881 (399)	0.723	0.471
Population	1,150,642 (168,114)	979,142 (180,925)	1.37	0.173
N = 153 border clusters				
Means Above Standard Errors in Parentheses				

E.2

We might also worry that the border counties are systematically different from non-border counties in observable ways. This would hurt the generalizability of the results. As such, I compare border counties to interior counties in terms of the average number of doctors, the average number of non-federal doctors, the average population and the average income in Table 19. These comparisons are at the county level rather than at the border cluster as in the previous comparison, so there are many more observations.

Table 20: Tests for Observable Selection into Border Sample

variable	Border Sample	Non Border Sample	t-stat	p-value
Physicians	297.404 (864.491)	330.548 (1175.56)	0.792	0.428
Non-Fed Physicians	34.827 (81.697)	36.932 (99.333)	0.553	0.580
Income	24,203 (6249)	24,413 (6201)	0.763	0.446
Population	127,310 (287,262)	123,482 (346,498)	0.286	0.774
N=664 border, N=2221 non-border				
Means Above Standard Errors in Parentheses				

Appendix F - Framework for Simulating The Size of the Incentive Effects

F.1 Supply

Firm free riding may be an optimal strategy in a game with positive spillovers. To investigate the incentives generated by the demand problem above, we would like to measure the marginal revenues generated by this type of advertising

and compare those with the marginal costs. Since advertising has a dynamic effect through carry-over, one must incorporate the dynamics through solving a dynamic programming problem. I should note here that the purpose of this exercise is not to predict firm behavior but to illustrate the size of the incentive effects of advertising spillovers. To do this, I will assume a very stylized model where firms optimize profits only over advertising and are forced to hold pricing and detailing fixed.

I assume that firms play a simultaneous game, choosing advertising each period while taking into account expectations of rival behavior and the dynamic effects of advertising. This enables me to analyze the magnitude of potential under-provision levels on average generated by positive spillovers in a very stylized game where only DTC advertising is moving.

F.1.1 The Firm's Problem

A forward looking firm maximizes a discounted stream of future profits with respect to advertising. Suppose advertising has a constant marginal cost k_{jm} , the market size is μ_m , prices are p_j , marginal production costs are mc_j and the discount rate is β . Further, suppose the exogenously given set of products that a firm f has in the market is denoted by Φ_f and the full set of products in the market is denoted by $\cup_f \Phi_f$.

Per period profit for the firm will be a function of advertising stock at the product, subcategory and category level $\mathbf{A}_{mt} = \{A_{jmt}, A_{nmt}, A_{lmt}\}$, which is a function of the vector of current advertising for all products, \mathbf{a}_{mt} , advertising stocks in the previous period $\mathbf{A}_{m,t-1}$, and persistence parameters λ_j, λ_n , and λ_l . Per period profit is also a function of other variables X , a constant marginal cost of advertising k_{jmt} , and product-market-time specific iid demand shocks ξ :

$$\pi_{fjm} = \sum_{j \in \Phi_f} (p_{jt} - mc_{jt}) \mu_{mt} s_{jmt}(\mathbf{A}_{mt}, X, \xi) - k_{jmt} a_{jmt}.$$

The firm's problem is to maximize the stream of future profits for all products in its portfolio. Advertising is set prior to the realization of any demand shocks, so only expected profits matter for firm choices. Expected per period profits are:

$$\pi_{fjm} = \sum_{j \in \Phi_f} \int (p_{jt} - mc_{jt}) \mu_{mt} s_{jmt}(\mathbf{A}_{mt}, X, \xi) p(\xi) d\xi - k_{jmt} a_{jmt}. \quad (28)$$

The timing of the game is as follows. At the start of each period, the state of the market \mathbf{g}_{mt} , is revealed to all firms. Based on that state, firms will make advertising decisions $\sigma_{jm}(\mathbf{g}_{mt}) = a_{jmt}$. The state variables could include anything in the past history of the game, but I will restrict attention to games where payoff-relevant state variables are the only ones that matter. Given the demand system provided above, state variables will include only the current advertising stock $\mathbf{A}_{mt} = \{A_{jt}, A_{nmt}, A_{lmt}\}$. After the state variables are observed, the firm makes its advertising decision, the current period demand shocks are realized and the firms collect their profits for the period.

The strategy profile $\sigma = (\sigma_1, \dots, \sigma_J)$ contains the decision rules of the firms. Firm f 's expected present discounted value of all profits given the current state \mathbf{g}_{mt} and the full strategy profile σ is

$$V_{fm}(\mathbf{g}_{mt}|\sigma) = \mathbb{E}\left[\sum_{j \in \Phi_f} \sum_{\tau=t}^{\infty} \beta^{\tau-t} \pi_f(\mathbf{g}_{ms}, \sigma_f(\mathbf{g}_{ms})) | \mathbf{g}_{mt}\right]. \quad (29)$$

Firm f chooses a sequence of advertising levels σ_f , which are state dependent, to maximize $V_{fm}(\mathbf{g}_{mt}|\sigma)$. I will assume that firms form expectations regarding future states and thus future strategies given current period state variables. As such, firm f will have a *Markov strategy*, $\sigma_{fm} : \mathbf{g} \rightarrow \sigma_f(\mathbf{g}) = (a_{jm})_{\forall j \in \Phi_f}$. To assess the strategies of competitors, firm f makes an assumption about σ_{-f} and chooses σ_f .

The equilibrium concept used will be *Markov perfect equilibrium*. A *Markov perfect equilibrium* is a set of strategies that form a subgame perfect equilibrium where strategies may only be conditioned on payoff relevant state variables and on the current state of the game. In particular,

$$\sigma_f^* \in \operatorname{argmax}\{V_{fm}(\mathbf{g}_{mt}|\sigma)\} \quad (30)$$

for all states \mathbf{g}_{mt} , firms f , and actions a_{jm} . That is, each firm will maximize the discounted sum of expected profits given beliefs about competitors behavior, and each firm's beliefs are mutually consistent in equilibrium. Using MPE makes the allowable strategy space relatively simple. I will further restrict attention to pure strategies in advertising. Firms will use the Bellman equation to solve the dynamic programming problem:

$$V_j(\mathbf{g}|\sigma) = \sup_{a \in \mathbb{R}^+} \{\pi_j(\mathbf{g}, a, \sigma_{-j}(\mathbf{g})) + \beta \mathbb{E}V(\mathbf{g}'|\sigma)\} \quad (31)$$

F.2 MPE Simulation

I solve the Markov Perfect Equilibrium (MPE) of the advertising game using a policy iteration algorithm, similar to Dube et. al. 2005. I take an initial guess of the strategy profile $\sigma^0 = (\sigma_1^0, \dots, \sigma_J^0)$, and then take the following steps:

1. For the strategy profile σ^n , calculate the value functions for each of the J firms, V_j^n . The value functions are defined by the Bellman equation (31). In this step, the maximization on the right hand side is not carried out, but the current guess of σ^n is plugged in.
2. For all $n > 0$, check whether the value functions and policy functions satisfy pre-determined convergence criteria, $\|V_j^n - V_j^{n-1}\| < \varepsilon_V$ and $\|\sigma_j^n - \sigma_j^{n-1}\| < \varepsilon_\sigma$. If so, stop the process.
3. Update each firm's strategy using the Bellman equation (31). Do this by carrying out the optimization on the right hand side. Denote the resulting policies and value functions by σ^{n+1} and V^{n+1} . Return to step 1.

The value functions are represented on a grid. As in the supply simulation, there are two firms, there are two firm advertising stocks and a category advertising stock. That makes three state variables. The grid is 5x5x5. Outside of the grid, the value function is determined using bilinear interpolation.

For the two-firm co-operative counterfactual, the MPE simulation turned into basically a monopoly game simulation where both firms' profits are considered and no other firms advertise. For the full co-operative counterfactual, computation is the same, only the advertising firm considers the profits of all products on the market.

Using the derived policies in each of these computations, the equilibria are simulated forward given the market initial conditions to generate the figures.

Available at Request - LDV Simulation

As there is a concern that a lagged dependent variable (LDV) will have a biased estimate with fixed effects in a panel model due to correlation between the lagged dependent variable and the error term, the purpose of this appendix is to simulate the size of such a bias. The theoretical size of the bias is computed in the text, but that only holds as $N \rightarrow \infty$. The finite sample simulation will assess how good of an approximation that is. In particular, I assume that the data generating process is:

- $\lambda_{category} = 0.68$
- $\alpha_{category} = 0.0494$
- $\lambda_{subcategory} = 0.28$
- $\alpha_{subcategory} = 0.00662$
- $\lambda_{product} = 0.324$
- $\alpha_{product} = 0.025$

With product-border-market, product-border-time, subcategory-border-market, subcategory-border-time, category-border-market and category-border-time fixed effects drawn from random normal distributions initially, all mean zero and the market effects with standard deviation of 1 and time effects standard deviations of 0.5.

Using 500 draws of epsilons with standard deviations that match the residual standard deviations of product, subcategory and category dependent variables, I simulate data, estimate the parameters using the same methods described in the paper. I then compute the difference between the estimated parameters and the true values that I set, $\hat{\lambda}_l - \lambda_l$ and $\hat{\alpha}_l - \alpha_l$. The results of the simulation are as follows:

Table 21: Simulated LDV Bias

Variable	Mean	SD
$(\hat{\lambda}_{category} - \lambda_{category})$	-0.0208	0.00258
$(\hat{\alpha}_{category} - \alpha_{category})$	-0.000790	0.00598
$(\hat{\lambda}_{subcategory} - \lambda_{subcategory})$	-0.0401	0.00171
$(\hat{\alpha}_{subcategory} - \alpha_{subcategory})$	0.00132	0.00933
$(\hat{\lambda}_{product} - \lambda_{product})$	-0.0223	0.000911
$(\hat{\alpha}_{product} - \alpha_{product})$	0.00009	0.00450