

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

Bradley T. Shapiro\*

This Version September 12, 2013

## 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 policy change that makes such advertising possible in the United States along with a spatial identification at the border approach, 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. Using these estimates and a simple supply model, I explore the consequences of the positive spillovers on firm advertising choice. In a cooperative advertising campaign, simulations suggest that the co-operative would produce on average four times as much advertising as is observed in competitive equilibrium, resulting in a 11.6 percent increase in category size and a 16.5 percent increase in category profits.

---

\*Massachusetts Institute of Technology. I thank Nancy Rose, Ernst Berndt, and Stephen Ryan for a wealth of advice throughout this project. I would also like to extend my gratitude to Sara Fisher Ellison, Glenn Ellison, Greg Leiserson, Sarah Moshary and Xiao Yu May Wang for their helpful suggestions. This paper also benefited from conversations with Isaiah Andrews, Dan Barron, Brad Larsen, Conrad Miller, Michael Powell, Adam Sacarny, Annalisa Scognamiglio, Paul Snyderman, Henry Swift and Len Tacconi; as well as participants at the MIT Industrial Organization field lunch. I thank IMS Health and Kantar Media 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.

# 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 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. To quantify the effect 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 the realized advertising outcome in the antidepressant market.

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 predated 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 is still 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 whether advertising does exhibit positive spillovers onto rivals. 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. Given the demand estimates and a model of supply, I back out the marginal costs of advertising from a model of dynamic strategic firm interaction. Finally, given estimates of the demand effects and marginal costs of advertising, I quantify the importance of free riding by re-simulating the supply model, assuming that a co-operative sets advertising for the entire industry.

Studies of Direct-to-Consumer (DTC) advertising in pharmaceuticals with varying credibility of identification strategies have suggested the possibility that cross-advertising elasticities could be positive, but results have been mixed. 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 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. In work that is more closely related to this study, 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. 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. Those that model both rely solely on time series data. This paper will model the full decision process and use data with both spatial and time series variation.

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 the magnitude of the potential positive spillover effects on advertising expenditure decisions.

Outside of the pharmaceutical literature, Sahni (2013) finds experimental evidence of positive spillovers to rivals in restaurant advertising in India. Additionally, Lewis & Nguyen (2012) and Anderson & Simester (2013) find evidence of positive spillovers in a number of categories.

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 in pharmaceuticals 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. In addition, I exploit a 1999 policy change making television advertising possible in the United States. I show that advertising has significant positive effects on rivals' sales, though smaller than its effects on own firm sales. Next, 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 some inter-temporal influence. Results indicate that advertising of antidepressants affects all levels of choice. The category effects are larger and more persistent over time than are business stealing effects, leading to a net positive spillover. Finally, I conduct a supply side analysis to evaluate to what extent positive spillovers suppress the incentive to advertise. I find that a co-operative deciding all advertising expenditure levels would advertise on average four times as much as is observed in competitive equilibrium. 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 DMA location of a county determines which local television stations that a consumer of cable or satellite dish gets with his or her subscription. In addition, those who watch television over the air, are more likely to pick up stations within their DMAs than from others. There are 210 DMAs in the United States, the largest 101 of which are included in my data.

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 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.

## 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.<sup>1</sup> Table 1 provides some 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. The

---

<sup>1</sup>A 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.

pattern is very similar in number of commercials, indicating that expenditures per capita are a reasonably comparable object across localities and national. However, as commercials are likely to be priced by reach, using expenditures per 100 capita should be a reasonably comparable way to measure reach of the ads. Figure 4 shows that national commercials and national expenditures per 100 capita are highly correlated.

### 2.2.3 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 prescriptions 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. In figure 5, the quantity weighted average margin in Boston is plotted versus time for the purposes of illustration. The average margin in the market rises as more popular branded antidepressants replace old generic TCAs and falls as more newer generation generics become widely prescribed.

Yearly county population and income data are drawn from the Current Population Survey (CPS).

Notable in the data is that there is both national and local advertising. While national advertising makes up the majority of advertising, there is significant spatial variation in the local additions to what households see.

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, Wellburin 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).

Finally, own firm and rival firm advertising are negatively correlated, which could be an indicator of the positive spillovers of advertising in this setting.

### 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,  $a^{cross}$ :

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

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 identify the effects of television advertising 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.

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 6.

To get an idea of how advertising is distributed across the country, consider the example of the Cleveland and Columbus DMAs. Figure 7 depicts the state of Ohio with each DMA in a different color<sup>2</sup>. Every county in the Cleveland, Ohio DMA gets the same amount of the same advertising as every other county in the Cleveland DMA. Meanwhile, every county in the 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 8. 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.

To estimate the effects of advertising in this experiment, I will use a modified difference-in-differences estimator. My identification 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

---

<sup>2</sup>from <http://www.dishuser.com/TVMarkets/Maps/ohio.gif>

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  $f$ :

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

Each product-border pair will constitute an experiment with border-markets being treatment groups. My fixed effects specification is:

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

where the subscripts  $j$  and  $b$  indicate which experiment is being considered (product and border specific),  $\alpha_{jbq}$  is a time effect which is used to control the experiment, which in this case will be a quarter fixed effect,  $\alpha_{jbm}$  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.

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,  $\alpha_{jbq}$ , 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,  $\alpha_{jbm}$ , 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. Since prior to 1997, no DMAs had any advertising, there will be at least some variation in each experiment over time.

### 3.1.1 Potential Threats to the Border Strategy

One potential worry is that there would be little variation after partialling out the fixed effects. This would be the case if too much of the advertising were national and not enough were local. Figure (9) 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 even after fixed effects are partialled out.

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.

Additionally, we might be concerned about measurement error. There are a two 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. Consumer 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.

Both of these scenarios would lead me to understate the effect of the advertisements. To the extent that we think that these biases are present, we can look at my estimates as lower bounds on the true parameters.<sup>3</sup>

Omitted variables bias could also be a source of bias. Prices 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. Detailing is observed in my data, but only at the national product-month level. Any national average effects of detailing are also controlled for

---

<sup>3</sup>However, 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.

by the product-border-time effect. Where there might be trouble, however, is if firms strategically raise (or lower) detailing at the product-market-time level in exactly the same places where DTC to take advantage of any complementarities or substitutabilities.

Two points help address the concern of omitted variables. First, using the border strategy, for such detailing to be a problem, firms would have to instruct detailers to stop detailing increases (decreases) exactly at the borders of the television market areas, which seems unlikely. Next, there is a literature (Manchanda et. al. 2004) suggesting that detailing is largely determined by practice size, which is effectively controlled for in the treatment group fixed effects. Finally, national detailing is much more stable over time than is national television advertising. As a robustness check, I have taken the national detailing data and assumed it was distributed across DMAs in exactly the same proportions as DTC advertising. This attribution of detailing constitutes a ‘worst case scenario’ in terms of how much detailing would bias any estimates of the effects of DTC. Doing this does not significantly change the results. This robustness check is provided in the appendix.

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.

### **3.1.2 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), henceforth BLP. 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. I am more comfortable with assumptions required by the border strategy in addition to the fact that it allows me to use the spatial variation in the data.

## 3.2 Results

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

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

where  $\alpha_{jbq}$  is a product-border-quarter fixed effect and  $\alpha_{jbd}$  is a product-border-DMA fixed effect. The  $\alpha_{jbq}$  effect will sweep out all variation that is not between two areas that are on opposite sides of a DMA border.

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. 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. If the category expansive effect of advertising is more persistent, it will be another potential avenue for firms to under-invest relative to a co-operative.

## 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).

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

$$u_{iImt} = \Gamma_1(A_{Imt}) + \beta_1 X_{Imt} + \xi_{It} + \xi_{Im} + \epsilon_{iImt} = \delta_I + \epsilon_{iImt}. \quad (3)$$

In this specification,  $I$  denotes the inside option versus outside option,  $m$  denotes market and  $t$  denotes time period. I define  $\Gamma_1$  as an increasing function of  $A_{Imt}$ , total advertising stock of all inside option products in market  $m$  at time  $t$ ,  $\xi_{It}$  is a time specific taste for the inside option,  $\xi_{Im}$  is a market specific taste for the inside option, and  $X_{Imt}$  are market-time characteristics.

For the next stage, I define the utility  $v^I$  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:

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

Finally, 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:

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

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  $l$ ,  $a_l$ , where  $l \in \{I, n, j\}$ , last period's advertising stock,  $A_{lm,t-1}$  and a parameter governing depreciation over time,  $\lambda_l$ .

$$A_{lmt} = f(\lambda_l, A_{lm,t-1}, a_{lmt}) \quad (6)$$

I set each disturbance term,  $\epsilon$ , 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:

$$s_j = (s_{j|n})(s_{n|I})(s_I) \quad (7)$$

Those conditional shares take logit form:

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

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

$$s_I = \frac{\exp(\delta_I)}{1 + \exp(\delta_I)} \quad (10)$$

I note here that the equations at each level are independent of each other. I allow each level to have a different persistence  $\lambda_l$ , and different effects of advertising,  $\gamma$ . 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.

For intuition, 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  $a_{Imt}$ , which increases  $A_{Imt}$  which then in turn increases  $\Gamma_1(A_{Imt})$ . 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.

#### 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:

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

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

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

and advertising elasticities equal to,

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

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  $s_j \frac{\partial \Gamma_1}{\partial a_j} (1 - s_I)$ , through expansion of the subcategory in  $s_j \frac{\partial \Gamma_2}{\partial a_j} (1 - s_{n|I})$  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  $s_j \frac{\partial \Gamma_1}{\partial a_k} (1 - s_I)$  or through expansion of the subcategory in  $s_j \frac{\partial \Gamma_2}{\partial a_k} (1 - s_{n|I})$ , 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. This is a testable implication of the model.

## 4.2 Supply

Firm free riding may be an optimal strategy in a game with positive spillovers. Mixed strategy equilibria may be an equilibrium in a specific game with a fixed cost to advertising each period. To investigate the incentives generated by the demand problem above, 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.

### 4.2.1 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  $\mu_m$ , prices are  $p_j$ , marginal production costs are  $mc_j$  and the discount rate is  $\beta$ . Further, suppose the exogenously given set of products that a firm  $f$  has in the market is denoted by  $\Phi_f$  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 of all products,  $\{A_{jmt}\}_{j \in \Phi}$ , captured in the vector  $\mathbf{A}_{mt}$ , which is a function of the vector of current advertising for all products,  $\mathbf{a}_{mt}$ , advertising stocks in the previous period  $\mathbf{A}_{m,t-1}$ , the persistence parameters,  $\lambda_j, \lambda_n$ , and  $\lambda_I$ , and the product-market-time specific constant marginal cost of advertising,  $k_{jmt}$ :

$$\pi_{fmt} = \sum_{j \in \Phi_f} (p_{jt} - mc_{jt}) \mu_{mt} s_{jmt}(\mathbf{A}_{mt} | \mathbf{a}_{mt}, \mathbf{A}_{m,t-1}, \lambda) - k_{jmt} a_{jmt} \quad (14)$$

The firm's problem is to maximize the stream of future profits for all products in its portfolio:

$$\max_{\{a_{jt}\}_{j \in \Phi_f}} \sum_{j \in \Phi_f} \beta^{\tau-t} \Sigma_j (p_{jt} - mc_{jt}) \mu_{mt} s_{jmt}(\mathbf{A}_{mt} | \mathbf{a}_{mt}, \mathbf{A}_{m,t-1}, \lambda) - k_{jmt} a_{jmt} \quad (15)$$

Which may be written as the recursive function,

$$V(\mathbf{A}_m | \mathbf{a}_m, \mathbf{A}_m^-) = \max_{\{a_{jm}\}_{j \in \Phi_f}} \{ \pi_m(\mathbf{A}_m | \mathbf{a}_m, \mathbf{A}_m^-) + \beta V(\mathbf{A}_m^+ | \mathbf{a}_m, \mathbf{A}_m) \} \quad (16)$$

As long as the composition of the advertising stock function with the response function  $\Gamma$  are concave in advertising, the problem has a well behaved optimum.

### 4.2.2 Other Choice Variables

Other choice variables are excluded from my model for reasons of expositional clarity and data limitation. Prices are observed only at the product-quarter level and detailing only at the product-month (national aggregate) level. While it is indeed conceivable that firms would maximize profit jointly with respect to both price and advertising, the purpose of this study is to isolate the advertising decision. In many settings

this may not be possible. However, in the pharmaceutical market that I study, the institutional detail of insurance as an intermediary makes pricing largely decided by a collection of bilateral negotiations between the firm, insurance companies and health care providers. The resulting prices observed in the data are highly persistent over time and show no correlation with observed advertising. As such, I will view the pricing decision as orthogonal to the DTC decision. Under this assumption, there will be no bias in the estimation of the effects of DTC rising from omitting price from the estimation. A discussion of pricing in this market as well as the analysis showing that it is not correlated with advertising is available in Appendix A.<sup>4</sup>

Detail advertising to the physician is another choice variable and type of marketing pursued in the pharmaceutical world. It is observed in the data, but only in national aggregates for each product over time. As with prices, detailing will be considered an orthogonal decision to DTC. Is this plausible? Conversations with physicians seem to indicate that they are being detailed about as much as they possibly can. Even if a firm wants to greatly ramp up detailing, it is unlikely that they will be able to if for no other reason than the physician only has so many hours in the day. If physicians are already satiated with detailing, this equilibrium effect would not be present. In addition, Manchanda et. al. (2004) find that high volume prescribers are detailed more than low volume prescribers without regard to their responsiveness to detailing.

Omission of pricing and detailing may still affect some conclusions in counterfactual simulations and estimations and discussion of those concerns will be explored in those sections.

## 5 Empirical Specification and Estimation of the Model

### 5.1 Demand Specification

I define the advertising stock at each level  $l$ , where  $l \in \{I, n, j\}$  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).

$$A_{lmt} = \sum_{\tau=0}^t \lambda_l^{t-\tau} \log(1 + a_{lm\tau}) \quad (17)$$

---

<sup>4</sup>Without including price in the analysis, I will not be estimating a price elasticity. As such, the computation of optimal advertising will be a bit different in spirit than the classic Dorfman-Steiner (1954) problem which suggests that the advertising to sales ratio should equal to the advertising elasticity to price elasticity ratio. This theorem was modified to allow for dynamic advertising in Nerlove and Arrow (1962). Neither of these formulations considers positive spillover effects, and free riding incentives may make these policies sub-optimal.

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.

$$\Gamma_l(A_{lmt}) = \gamma_l A_{lmt} \quad (18)$$

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 (17) and (18) into equations (1)-(3), for the market level I obtain:

$$u_{iImt} = \gamma_l [\sum_{\tau=0}^t \lambda_I^{t-\tau} \log(1 + a_{Im\tau})] + \xi_{It} + \xi_{Im} + \epsilon_{iImt} \quad (19)$$

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 will be defined as the population not filling a prescription for an antidepressant in month  $t$  in market  $m$ :

$$\log(s_{Imt}) - \log(s_{omt}) = \gamma_1 [\sum_{\tau=0}^t \lambda_I^{t-\tau} \log(1 + a_{Im\tau})] + \xi_{It} + \xi_{Im} \quad (20)$$

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:

$$\log(s_{nmt|I}) - \log(s_{omt|I}) = \gamma_2[\sum_{\tau=0}^t \lambda_n^{t-\tau} \log(1 + a_{nm\tau})] + \xi_{nt} + \xi_{nm} \quad (21)$$

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:

$$\log(s_{jmt|n}) - \log(s_{omt|n}) = \gamma_3[\sum_{\tau=0}^t \lambda_p^{t-\tau} \log(1 + a_{jm\tau})] + \xi_{jt} + \xi_{jm} \quad (22)$$

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

$$\log(s_{Imt}) - \log(s_{0mt}) = \lambda_I(\log(s_{Im,t-1}) - \log(s_{0m,t-1})) + \gamma_1 \log(1 + a_{Imt}) + \theta_{It} + \theta_{Im} \quad (23)$$

where

$$\theta_{It} = \xi_{It} - \lambda_I \xi_{I,t-1} \quad (24)$$

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

and

$$\theta_{Im} = \xi_{Im} - \lambda_I \xi_{Im} \quad (25)$$

is the category-market specific taste parameter. This 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:

$$\log(s_{nmt|I}) - \log(s_{0mt|I}) = \lambda_n(\log(s_{nm,t-1|I}) - \log(s_{0m,t-1|I})) + \gamma_2 \log(1 + a_{nmt}) + \theta_{nt} + \theta_{nm} \quad (26)$$

$$\log(s_{jmt|n}) - \log(s_{0mt|n}) = \lambda_p(\log(s_{jm,t-1|n}) - \log(s_{0m,t-1|n})) + \gamma_3 \log(1 + a_{jmt}) + \theta_{jt} + \theta_{jm} \quad (27)$$

### 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:

$$\log(s_{I_{mt}}) - \log(s_{0_{mt}}) = \lambda_I(\log(s_{I_{m,t-1}}) - \log(s_{0_{m,t-1}})) + \gamma_1 \log(1 + a_{I_{mt}}) + \alpha_{bt} + \alpha_{bm} + \epsilon_{mt} \quad (28)$$

where  $\alpha_{bt}$  is a border-time fixed effect and  $\alpha_{bm}$  is a border, DMA fixed effect. Partialling 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 time period  $t$  in all counties on that either side of border  $b$ . The fixed effects will also control for the product quality terms  $\theta_t$  and  $\theta_m$ .

I identify the effects at the other two levels similarly. In the subcategory level, I include fixed effects  $\alpha_{nbt}$  and  $\alpha_{nbm}$  and at the product level, I include fixed effects  $\alpha_{jbt}$  and  $\alpha_{jbm}$ . 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 time  $t$  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 4. 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. Table 5 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- that is getting someone into consuming antidepressants at all. The persistence parameter of 0.68 implies that 90% of the effect dissipates within six months. Meanwhile, the 0.33 persistence parameter on the bottom 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 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.

## **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 through the incentivize to free riding on the efforts of rivals.

### 6.1.1 Internalization

To further illustrate the effects of advertising over time on rivals, consider an impulse response graph in figure (10). 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 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 would lead to a contemporaneous increase of about 70,000 antidepressant commercials, only about 20,000 of which would be Zoloft. Further, as we follow that effect 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

To illustrate the free riding incentive, I consider three marginal revenue curves and a horizontal marginal cost curve in figure (11) 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.

In the next subsection, I will more systematically explore these incentives for our realized antidepressant market.

## 6.2 Cost Computation

Given the demand estimation above, I can solve for the marginal costs associated with advertising implied by the outlined supply model. In spirit, this means that I am assuming that firms are behaving optimally, computing marginal revenues in each market in each month for each product given the expected behavior of rivals, demand estimates and price and production cost and setting that equal to marginal costs. While at first blush, it may seem as though the marginal cost of one dollar of advertising should be one dollar, it is possible that there are opportunity costs to firms to advertising. They could have other product classes to advertise and limited advertising budgets set within the organization. They could have concerns about the public or regulator perception of advertising drugs too much (Ellison and Wolfram 2008). As in Ellison and Ellison (2011), they may be strategically trying to deter entry by making the market seem small. In principle, they could also have other marginal benefits to advertising. There may be cross class spillovers. As such, I will allow the ‘marginal cost’ of advertising one dollar to be either more or less than one dollar. Since not all firms advertise in all markets and in all months, I will only be able to bound costs below where firms do not advertise.

Summaries of estimated marginal costs are presented in table 6 for each product.

While the marginal cost of one dollar of advertising is consistently more than one, typically centered around four, it varies across product and market. These costs will be used to compute the counterfactual of co-operative advertising.

### 6.2.1 Pulsing Strategies

As is noted by Dube et. al. (2005) and is evident from the figures of observed advertising, firm advertising is often highly variable and unpredictable. While the model of Dube et. al. rationalizes the spikes and zeros in advertising with an advertising stock function that is s-shaped in current advertising advertising, I rationalize those spikes and zeros by changes in costs over time. Firms might plausibly have varying opportunity costs over time and markets due to portfolio changes or organizational concerns.

It is possible that the supply side game could generate equilibria with entry into advertising in mixed strategies, if for example there were fixed costs to advertising each period, which are not specified in this model. However, as the explanation of pulsing is not the main focus of this paper, details of such possibilities will be left to future research.

### 6.3 Counterfactual Assumptions

Given the positive spillovers of advertising, we should expect that the incentive to invest in advertising is lessened by both a failure to internalize the benefits of advertising on rivals and by an incentive to free ride. As such, I will consider a counterfactual scenario whereby the entire market is allowed to set advertising in a single optimization problem, thereby taking away strategic considerations. For such a scenario to work, cooperation would need to be not only allowed, but enforced in some way. Co-operatives in the milk, orange juice and beef industries were set up by state or federal governments to allow producers to sign enforceable contracts. Antidepressants might be able to cooperate similarly through non-profit organizations called patient advocacy groups. Such groups are focused on educated patients on specific diseases and treatments. While they do not tend to advertise on television currently, they might be an ideal facilitator for category level advertising of antidepressants.

For the purposes of the counterfactual, 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. The effect of those advertisements is equal to the category level effect of the branded advertisements estimated above.

The co-operative solves equation (29) in each month and in each market, taking as given the computed marginal costs of advertising each product from the previous section. I assume that the marginal co-operative advertisement has cost equal to the average of the computed marginal costs of each product, weighted by the amount those products advertise in a given period. I also assume that the co-operative can forecast margins and populations in future periods.

$$V(\mathbf{A}_m | \mathbf{a}_m, \mathbf{A}^-_m, \lambda) = \max_{\{a_{jmt}\}_j} \{ \pi_{mt}(\mathbf{A}_m | \mathbf{a}_{mt}, \mathbf{A}^-_m, \lambda) + \beta V(\mathbf{A}^+_m | \mathbf{a}_m, \mathbf{A}_m, \lambda) \} \quad (29)$$

Notably missing from this firm problem is the ability for each firm to readjust pricing and detailing with the co-operative advertising decisions, as pricing and detailing are not included in the model. As mentioned previously, prices are uncorrelated with levels of DTC in the observed world. There is little reason to suspect that they would change in the counterfactual. This is because most price flexibility happens in negotiations with managed care organizations which were less prevalent in the time of this sample and those negotiations happen infrequently and in a disjointed way from the advertising decision.

Detailing, on the other hand, could be more troublesome. If detailing were mainly a business stealing device, as is suggested in Narayan et. al. (2004), it would be complementary to co-operative DTC. A firm would be happy for the co-operative to make the market large and it would then try to compete over brand share in detailing space. Anticipating this, the co-operative would shade down its advertising, anticipating that profits would be competed away with detailing.

If detailing were, alternatively, also mainly category expansive, it would be a substitute for DTC. In response to the DTC, firms would be glad to reduce detailing and let the co-operative foot the bill for market expansion. The co-operative would anticipate this and shade up its advertising. Conversations with physicians seem to indicate that they are being detailed about as much as they possibly can. Even if a firm wants to greatly ramp up detailing, it is unlikely that they will be able to if for no other reason than the physician only has so many hours in the day. If physicians are already satiated with detailing, the potential equilibrium effect of competing away all market expansion with further detailing would not be present.

Also worth noting, since in the observed world firms under invest in advertising, the counterfactual results in some values of advertising that are not observed in the sample. Since we only estimate the demand curve for values of advertising within the sample, the out-of-sample specific numbers are driven by the assumed functional form of the advertising effect in the model.

## 6.4 Advertising

As the business stealing incentive grows, observed 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 observed advertising relative to the co-operative's ideal.

I assume that the co-operative can set a number of non-branded co-operative advertisements that will have the market level effects and persistence estimated in the demand section. It will be able to do so at the average cost of advertising across all products in that market and period, with the restriction that the marginal cost of advertising one dollar may not go below one dollar plus the industry standard 15% agency fee. The co-operative assumes stationarity of the optimization problem and maximizes discounted future profits in each month. Generics are assumed to have zero margin.

For illustrative purposes Figure (12) shows the observed flow of advertising and the co-operative's choice in Boston and Figure (13) shows the observed flow versus co-operative choice on average across all DMAs in

the simplified co-operative scenario described above. Figures (14) and (15) are the same as Figures (12) and (13) except that they assume that the marginal cost of advertising one dollar must be exactly one dollar plus the 15% agency fee.

Notably, the co-operative maintains significantly higher advertising over time than is the observed outcome. On average, the co-operative advertises four times as much as the competitive industry. The positive spillovers of advertising seem to be generating a free riding problem for the industry. The exact numbers are coming from extrapolation on the functional form, as they are often out of sample from what is estimated. Also, where the difference between counterfactual and observed advertising is low, computed marginal costs are very high. Where the difference is large, the computed marginal costs are low. If we assume that true costs of advertising a dollar in the co-operative advertising world do not exceed unity, differences are persistently much higher, with counterfactual advertising on the order of ten times observed advertising.

## 6.5 Quantities and Profits

In the co-operative, the greater advertising leads to an increase in shares of the inside option by 11.6% and an increase in total industry profit of 16.5% on average. Those averages over time are plotted in figures (16) and (17).

## 6.6 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 the antidepressant market or prescription pharmaceutical markets more generally. First, many prescriptions are covered by insurance. While many people are getting prescribed and incurring minimal if any cost, the system is still paying out the marked price, which is quite high. It is possible that the new prescriptions are not justified by the societal cost. 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 the higher advertising. If this is the case, the welfare effect is also ambiguous. However, as prices are not correlated with advertising expenditures and detailing might well have reached satiation, this concern might not be very large.

The actual net social welfare effect, while interesting, is not identified in this study and is certainly worthy of further research. Category expansion is typically thought of as consumer welfare improving, as consumers

would not purchase something for which the costs exceeded the benefit. However, in health care, many levels of agency mask the costs and benefits of treatments. As such, I am remaining agnostic on consumer welfare.

## 7 Conclusions

Using data from the antidepressant market and an identification strategy taking advantage of both policy and spatial discontinuities, I find that television advertising has significant positive spillovers. I construct and estimate 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 commons problem whereby observed advertising is significantly lower than the optimal strategy that a co-operative would set if it controlled the entire market. A co-operative would set advertising five times as high as is observed in equilibrium and would increase industry shares by 11.6% and profits by 16.5%.

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.

### 7.1 External Validity to other Contexts

While the result may not be the same in other industries, it is not unreasonable to expect that other advertising could show some similar positive spillovers. The existence of advertising cartels in milk, juice and beef might well be evidence of such effects. While the parameters estimated in the antidepressant industry are unlikely to be relevant to other industries, the model developed in this paper could be applied in other industries to measure the extent of the positive spillovers in those contexts.

## References

- [1] 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)

- [2] 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.
- [3] Anderson, E., Simester, D., (2013). Advertising in a competitive market: the role of product standards, consumer learning and switching costs. Working Paper.
- [4] 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.
- [5] 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.
- [6] Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, 25(2), 242-262.
- [7] Busch, S., Golberstein, E., & Meara, E. (2011). The FDA and ABCs: The unintended consequences of antidepressant warnings on human capital. NBER Working Paper # 17426.
- [8] 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.
- [9] 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.
- [10] Dorfman, R., & Steiner, P. O. (1954). Optimal advertising and optimal quality. *The American Economic Review*, 44(5), 826-836.
- [11] Dubé, J. P., & Hitsch, G. J., & Manchanda, P. (2005). An empirical model of advertising dynamics. *Quantitative Marketing and Economics*, 3(2), 107-144.
- [12] 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.
- [13] 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.
- [14] Ellison, S. F., & Wolfram, C. (2006). Coordinating on lower prices: pharmaceutical pricing under political pressure. *The RAND Journal of Economics*, 37(2), 324-340.

- [15] Goeree, M. S. (2008). Limited information and advertising in the US personal computer industry. *Econometrica*, 76(5), 1017-1074.
- [16] Iizuka, T. (2004). What Explains the Use of Direct-to-Consumer Advertising of Prescription Drugs?. *The Journal of Industrial Economics*, 52(3), 349-379.
- [17] Iizuka, T., & Jin, G. Z. (2005). The effect of prescription drug advertising on doctor visits. *Journal of Economics & Management Strategy*, 14(3), 701-727.
- [18] Iizuka, T., Jin, G. Z. (2005). Direct to consumer advertising and prescription choice. Working Paper Available at SSRN 700921.
- [19] Jayawardhana, J. (2013). Direct-to-consumer advertising and consumer welfare. *International Journal of Industrial Organization*, 31(2), 164-180.
- [20] 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.
- [21] 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.
- [22] Lewis, R. A., & Rao, J. M. (2012). On the near impossibility of measuring advertising effectiveness. Yahoo Research Working paper.
- [23] Manchanda, P., Rossi, P., & Chintagunta, P. (2004). Response modeling with non-random marketing mix variables. *Journal of Marketing Research*, 41(4), 467-478.
- [24] Murphy, M. J., R. L. Cowan, and L. I. Sederer (2009): *Blueprints: Psychiatry*, Philadelphia, PA: Lippincott Williams Wilkins, 5 ed.
- [25] 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.
- [26] Nerlove, M., & Arrow, K. J. (1962). Optimal advertising policy under dynamic conditions. *Economica*, 29(114), 129-142.
- [27] 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.

- [28] Sahni, N. (2013). Advertising Spillovers: Field Experimental Evidence and Implications for the Advertising Sales-response Curve. Working Paper.
- [29] 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).
- [30] 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

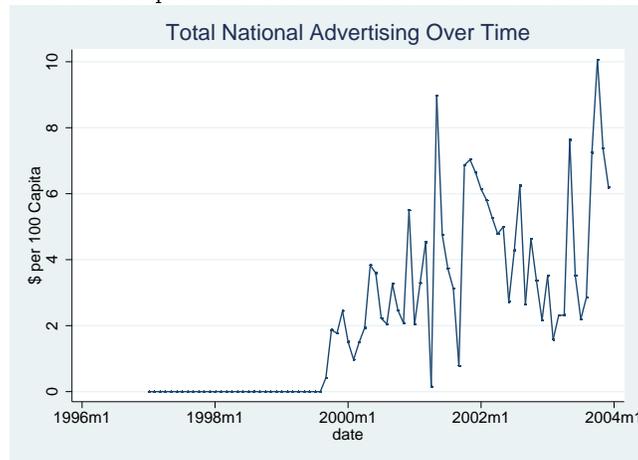


Figure 2: Antidepressant Revenues 1996-2008

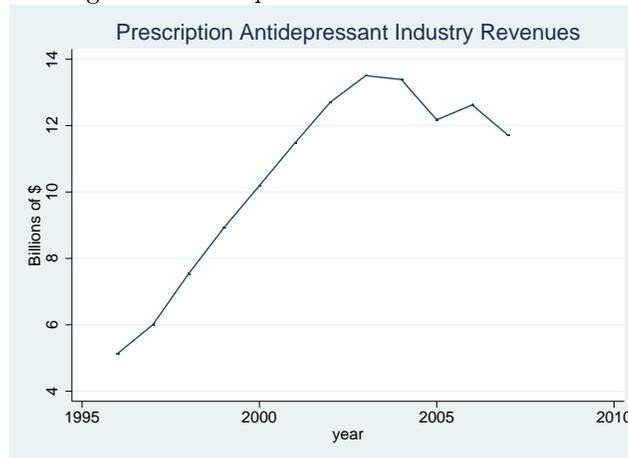


Table 1: Descriptive Statistics: Advertising

	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
DMAs	101			
DMA Population	2340774	903090	1469823	2622567

Figure 3: Variation Across Three Markets in Advertising

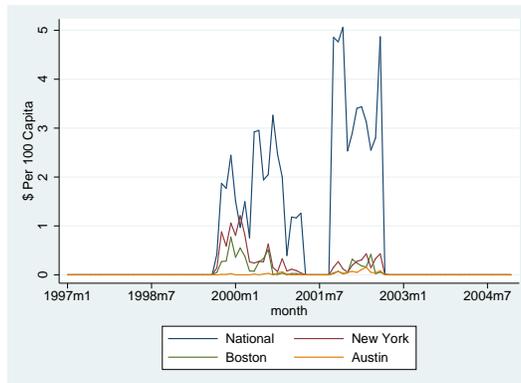


Figure 4: Two Different Advertising Measures: National

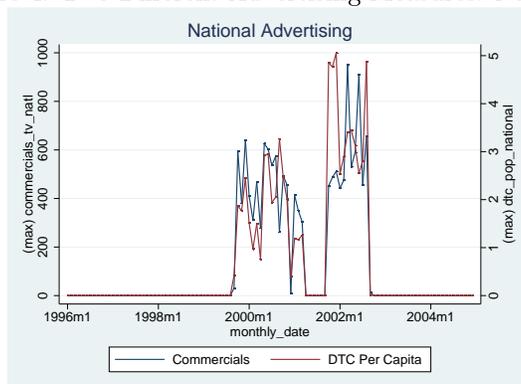


Figure 5: Quantity Weighted Average Margins in Boston

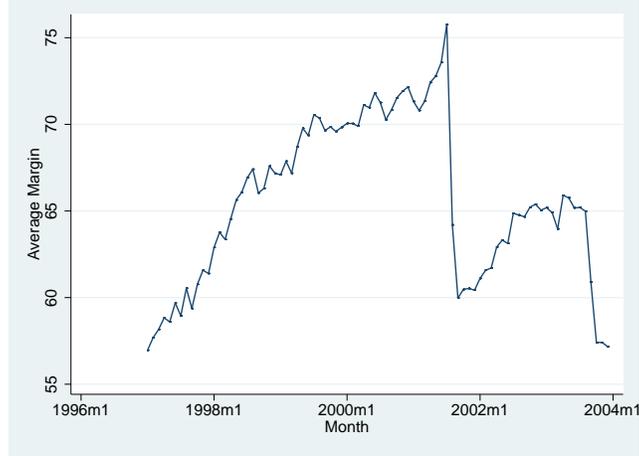


Figure 6: Full Sample: Top 101 DMAs

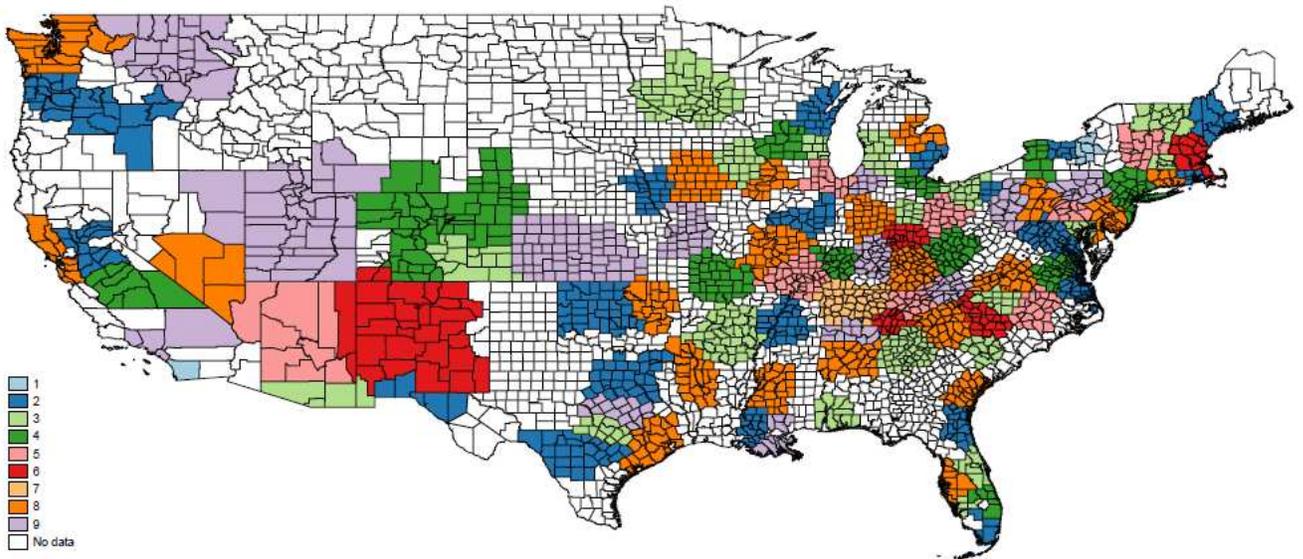




Figure 9: Variation in Log DTC Net of Fixed Effects, 14% Zeros

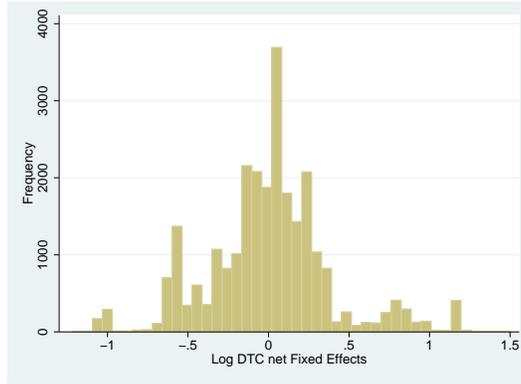


Table 2: The Effect of Own and Rival Advertisements on Sales

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

DMA clustered standard errors in parentheses

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

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

Figure 10: Impulse Response Effect of Zoloft Advertisement on Own and Total Prescriptions

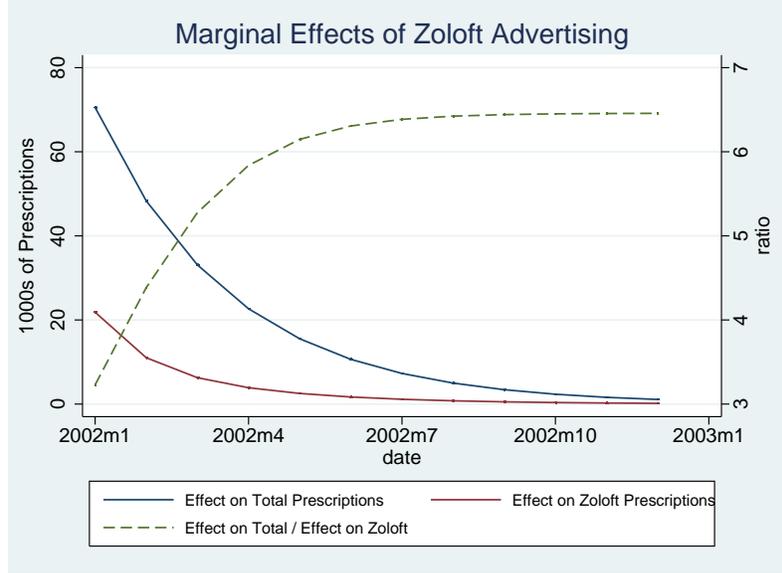


Table 4: Results of Base Model

VARIABLES	Category Level	Subcategory Level	Product Level
<i>adstock</i>	0.0472*** (0.00665)	0.0093 (0.00719)	0.0223*** (0.00756)
<i>persistence, λ</i>	0.684*** (0.0299)	0.282*** (0.0116)	0.330*** (0.0140)
Observations	23,091	93,284	60,980
R-squared	0.948	0.935	0.960

DMA clustered standard errors in parentheses

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

Table 5: 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 11: Marginal Revenue Curves Under Various Scenarios  
 Marginal Revenue of Advertising for Zoloft

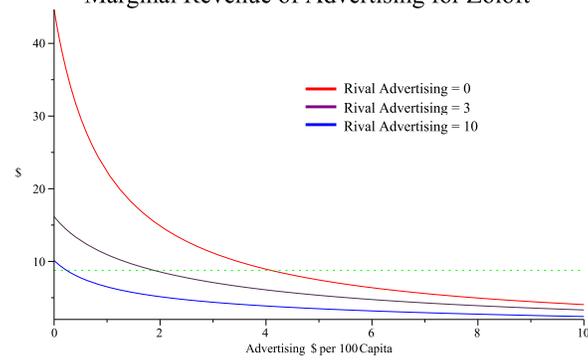


Table 6: Marginal Cost Distributions By Product

	Mean	Q10	Median	Q90
Paxil	3.993	2.098	3.611	6.334
Paxil CR	3.289	0.956	2.949	6.271
Prozac	4.014	1.433	3.782	6.844
Prozac Weekly	1.035	0.329	0.930	1.988
Wellbutrin SR	4.589	1.874	3.067	11.730
Wellbutrin XL	1.722	0.706	1.435	3.144
Zoloft	2.870	1.314	2.500	4.952

Figure 12: Counterfactual versus Realized Advertising in Boston

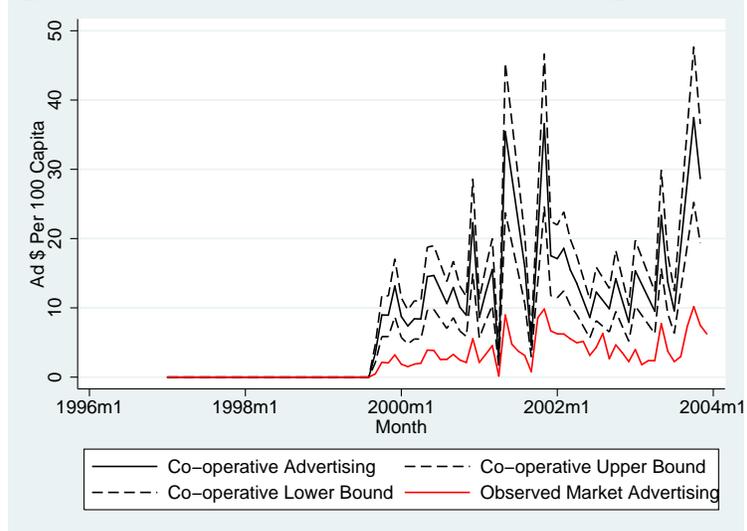


Figure 13: Counterfactual versus Realized Advertising on Average

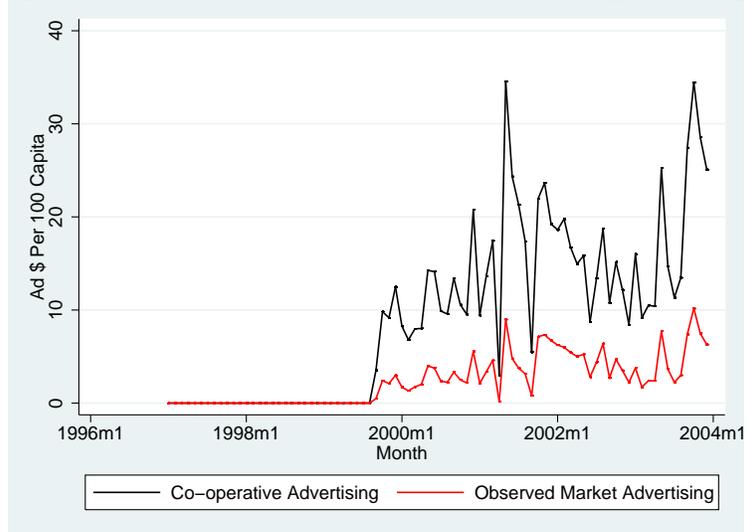


Figure 14: Counterfactual versus Realized Advertising in Boston, MC=\$1

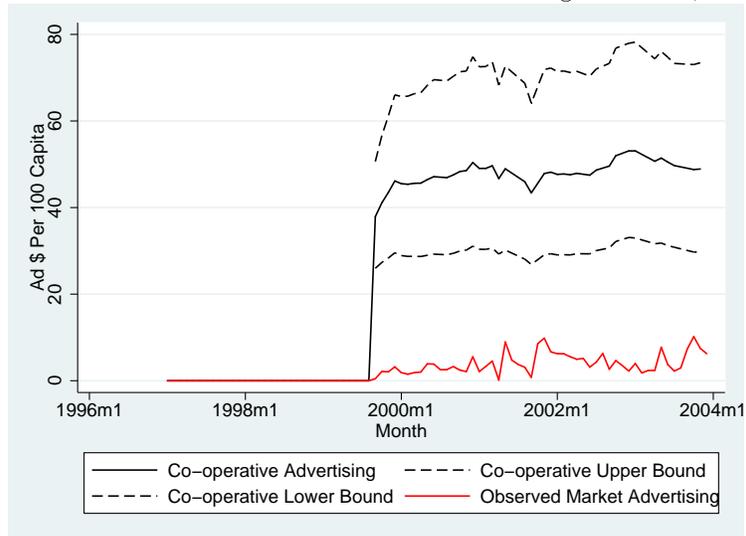


Figure 15: Counterfactual versus Realized Advertising on Average, MC=\$1

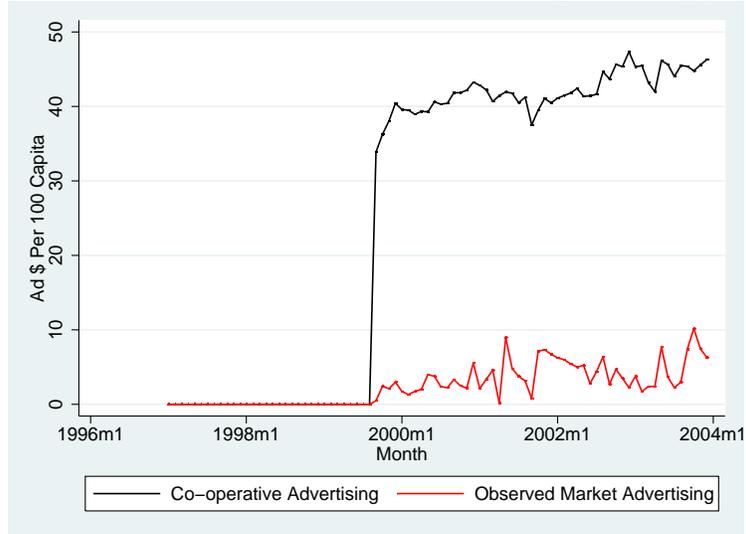


Figure 16: Counterfactual versus Realized Antidepressant Share Prescribed on Average

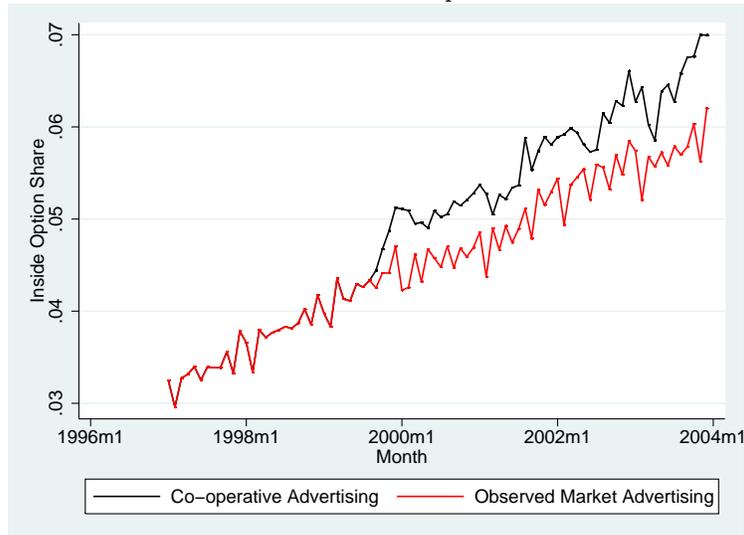
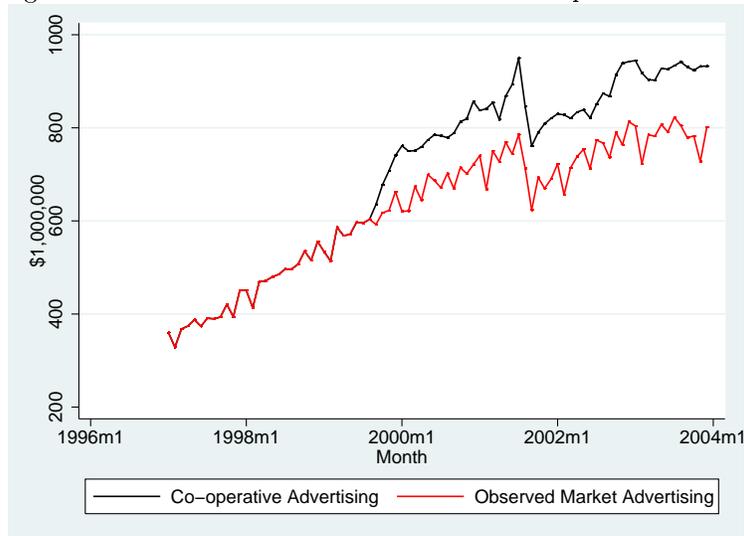


Figure 17: Counterfactual versus Realized Antidepressant Profits



## Appendix A - Orthogonality of Pricing and Detailing Decisions

### A.1 Pricing

Below is the regression of television commercials on prices. As can be seen the point estimate is very small and insignificant with respect to both own and cross advertising. Just for perspective, the average unit price of a branded drug is about \$3.60 over the course of the sample and the average DTC per capita of those

Table 7: Predicting Prices with Advertising

VARIABLES	(1) realunitprice
<i>DTC</i>	-0.0518 (0.0437)
<i>DTC_other</i>	-0.0332 (0.0248)
time	0.00963*** (0.00313)
expired	-0.210 (0.149)
Product FEs	yes
Observations	1188
R-squared	0.994

Product Clustered Standard errors in parentheses  
\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

drugs that advertise over the course of the sample is \$0.0055. Raising DTC per capita by \$0.01 is associated with a price decrease of \$0.03. This is very small economically. Also interesting is that just a time trend, a product fixed effect and a dummy for patent expiration can explain prices with R squared bigger than 0.99. Prices seem quite sticky, especially relative to advertising in this market.

## A.2 Detailing

To address the potential omitted variables bias problem in detailing, I assumed that the observed national detailing was geographically distributed in exactly the same way as television advertising. That is, I computed a DMA fraction of television advertising as the expenditures per capita divided by the sum over all markets of the expenditures per capita. As detailing and television advertising are, in general, positively correlated, I multiplied that fraction by national detailing totals to get a ‘worst case scenario’ to see how much detailing could possibly affect the estimated effect of television advertising. If detailing and television advertising were negatively correlated, I would want to assume detailing was inversely distributed to this.

The results of this ‘worst case scenario’ are presented in Table 8. None of the estimates on television advertising are statistically distinguishable from the baseline estimation. I will exercise caution in interpreting the coefficients on detailing, as they are not the actual detailing numbers at different localities.

## 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

Table 8: Results of Model including Worst Case Detailing

VARIABLES	Category Level	Subcategory Level	Product Level
<i>adstock</i>	0.0450*** (0.00666)	0.0102 (0.00731)	0.0197*** (0.00791)
<i>logdetail</i>	0.0316*** (0.0117)	0.00313 (0.00812)	0.0124*** (0.00946)
<i>persistence</i>	0.684*** (0.0302)	0.280*** (0.0116)	0.327*** (0.0143)
Observations	23,070	93,280	147,443
R-squared	0.948	0.935	0.960

DMA clustered standard errors in parentheses

\*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1

Table 9: Results of Base Model with Placebo

VARIABLES	Category Level	Subcategory Level	Product Level
<i>adstock</i>	0.0471*** (0.00672)	0.0109 (0.00730)	0.0216*** (0.00764)
<i>placebo</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

DMA clustered standard errors in parentheses

\*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1

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

## 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

Table 10: Results of Base Model without Northeast Corridor and Other Urban Areas

VARIABLES	Category Level	Subcategory Level	Product Level
<i>adstock</i>	0.0513*** (0.00828)	0.0063 (0.00894)	0.0279*** (0.00945)
<i>persistence</i>	0.688*** (0.0322)	0.255*** (0.0129)	0.313*** (0.0160)
Observations	17,340	67,796	42,568
R-squared	0.946	0.927	0.955

DMA clustered standard errors in parentheses  
 \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Table 11: Results of Base Model with only Northeast Corridor and Other Urban Areas

VARIABLES	Category Level	Subcategory Level	Product Level
<i>adstock</i>	0.0345*** (0.00949)	0.0200** (0.00914)	0.0071*** (0.0111)
<i>persistence</i>	0.648*** (0.0676)	0.387*** (0.0215)	0.381*** (0.0285)
Observations	5,751	25,488	18,412
R-squared	0.963	0.956	0.972

DMA clustered standard errors in parentheses  
 \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

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.

### 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 12: Results of Base Model Without Using the Border Approach

VARIABLES	Category Level	Subcategory Level	Product Level
<i>adstock</i>	0.0436*** (0.00140)	0.0030 (0.00248)	0.0187*** (0.00438)
<i>persistence</i>	0.747*** (0.0209)	0.721*** (0.0124)	0.595*** (0.0116)
Observations	8,216	41,465	32,710
R-squared	0.976	0.984	0.984

DMA clustered standard errors in parentheses

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

## C.2 Are the Borders Special? Full Sample Estimation

Next, I assume that the border sample approach is not necessary and that advertising may be viewed as predetermined with respect to demand shocks. This might be plausible since a large fraction of television advertising in the pharmaceutical market is purchased in the upfront market. Using a difference-in-differences approach with a lagged dependent variable and a common trend for all DMAs, I estimate the model. The point estimates for the effects of advertising are not statistically different from those in the border approach. However, the persistence parameters are a bit larger. If we believe these persistence parameters over the ones estimated in the border approach, the spillover problem will be even larger than was estimated. I continue to prefer the border approach, as the assumptions of common trends among similar geographies is more plausible than a nationally common trend. In addition, it gives more modest estimates of the size of the spillover, working against, my main finding of spillover effects.

## 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.