Choice of Depression Treatment: Advertising Spillovers in a Model with Complementarity*

Bogdan Genchev†

Job Market Paper

November 11, 2019
Click here for the most recent version

Abstract

I study how antidepressant direct-to-consumer advertising affects demand for psychotherapy. I develop a discrete-choice demand model that allows for complementarity between products, advertising spillovers, and flexible unobserved preference heterogeneity. I estimate the model using panel data on depression treatment choices by patients and antidepressant advertising data. The results indicate that even though antidepressants and psychotherapy are substitutes, drug advertising increases demand for therapy through a spillover effect. Allowing for time-invariant and time-varying unobservables that can be correlated across products critically affects the estimated degree of complementarity and advertising elasticities.

*I am grateful to Julie Holland Mortimer, Michael Grubb, and Charles Murry for their guidance and support. I thank Hunt Allcott, Donald Cox, Michael Dickstein, Kurt Lavetti, Alexander MacKay, Sarah Moshary, Michael Ostrovsky, Theodore Papageorgiou, Mathias Reynaert, Tobias Salz, Brad Shapiro, Cailin Slattery, Richard Sweeney, and participants at the Boston College dissertation workshop and Student Applied Micro Seminar for their useful comments and suggestions. Medical and prescription drug claims data were generously provided through the Truven Health Analytics MarketScan® Dissertation Support Program. Truven Health was not involved in preparing the results in this paper in any way. I am solely responsible for any errors.

†Department of Economics, Boston College, genchev@bc.edu

1 Introduction

Direct-to-consumer (DTC) advertising of prescription drugs on television has been controversial since its deregulation in 1997. One of the various criticisms leveled against this type of promotion is that it overemphasizes the use of pharmaceutical products at the expense of alternative treatment options.[1] In the context of depression, psychologists have expressed concern over the declining use of psychotherapy in depression treatment and pointed to antidepressant DTC ads as one of the reasons.[2]

I study how antidepressant TV advertising affects demand for psychotherapy. There are two possible channels: complementarity or substitutability between treatment options and spillover or business-stealing effects of advertising. If drugs and therapy are complements because they work better together, an increase in drug ads will increase demand for therapy. However, it is also possible that antidepressant advertising has a spillover effect by encouraging individuals to seek treatment, which can boost demand for therapy even if the products are substitutes. To disentangle the channels through which advertising operates, I develop a discrete-choice model that captures the key features of demand: complementarity, spillovers, and unobserved heterogeneity in preferences.

The model presents two identification challenges: separating complementarity from unobserved correlated preferences and identifying advertising spillovers. The first challenge arises from the fact that a relatively large share of patients may choose to take drugs and therapy either because they like both options for reasons unrelated to how well the two work together (unobserved correlated preferences) or because the products provide a greater benefit when taken in combination (complementarity). In addition, time-varying health shocks that boost demand for all treatment options can also make it seem that patients tend to choose drugs and therapy together relatively more often and falsely suggest that they are complements.

To separately identify complementarity from unobserved correlated preferences, I use

[1] Ventola (2011) outlines the arguments for and against DTC advertising of prescription drugs.
[2] Nordbal (2010) discusses various possible causes of this trend.

two sources of identification. The first is individual-level panel data. The intuition is that if drugs and therapy are complements, individuals will buy them in combination relatively more often than either one alone. I combine this with health plan level variation in prices. If demand for drugs and demand for therapy move in the same direction in response to an exogenous change in the price of either one, the products are complements; if they move in opposite directions, they are substitutes.

Identifying advertising spillovers typically requires separating own and cross effects when two products are advertising at the same time: if demand for both products increases, this can be rationalized by strong effects of own advertising and weak business-stealing of rival advertising or weak own effects and weak spillovers of rival advertising. This necessitates having two instruments to separately identify own and cross effects.[3] In the case of depression treatment, however, therapists are typically sole proprietorships that do little mass advertising. Thus, the only products that are advertised on TV are antidepressants, which significantly simplifies the problem. To alleviate concerns about the endogeneity of drug advertising, I use a cost-based instrumental variable.

I use two main sources of data. Individual-level panel data of treatment choices, prices, and patient demographics come from the Truven Health Marketscan® Commercial Claims and Encounters Database for the period 2008–2010. The Kantar Media Ad$pender database provides monthly advertising expenditures and counts of national and local TV ads for antidepressants.

My model allows for complementarity as well as time-invariant and time-varying unobserved preferences that can be correlated across products. It extends Gentzkow (2007) by allowing for multiple markets and time periods and accommodating estimation of own and spillover advertising effects using linear instrumental variable methods. To estimate the model, I follow an approach that combines individual and market-level data similar to

[3]In general, one needs at least as many instruments as endogenous variables. I consider the effect of aggregate branded antidepressant advertising on demand for generic drugs, branded drugs, and therapy, and need a single instrumental variable. With more instruments, it is possible to study own and cross effects for each product. Shapiro (2018) and Sinkinson and Starc (2019) are two papers that successfully estimate business-stealing and spillover effects at the product level.

the one in Berry et al. (2004) (micro-BLP) and Goolsbee and Petrin (2004).[4] Estimation proceeds in two stages. The first estimates the parameters on individual-specific observables and the distributions of the unobservables using maximum simulated likelihood (MSL) while "concentrating out" bundle-market-time fixed effects.[5] I assume that the time-varying health shocks follow a Markov chain and use a technique from the literature on Hidden Markov Models (HMMs) to significantly simplify the calculation of the likelihood. I then use simulation to approximate integrals over the time-invariant unobserved preferences. Because advertising effects are not separately identified from the bundle-market-time fixed effects, the second stage of the estimation projects the recovered fixed effects on ads and market and time effects. To estimate the causal effect of advertising on demand for generic drugs, branded drugs, and psychotherapy, I use the average price of a 30-second TV ad spot, adjusted for ratings, as an instrument.

The results indicate that antidepressants and psychotherapy are substitutes on average, although there is substantial variation across time and space. The degree of substitutability increases slightly over time. In some markets, drugs and therapy are complements.

The fact that drugs and therapy are substitutes would imply that, in the absence of spillovers, an increase in antidepressant advertising decreases demand for psychotherapy. However, the results show that there are strong advertising spillovers that make the net effect of drug ads on demand for therapy positive: on average, a 10 percent increase in advertising leads to a 0.093 percent increase in demand for therapy. The same change in advertising also leads to a 0.050 percent positive spillover on demand for generics and a 0.076 increase in demand for branded drugs.

The conclusions depend critically on allowing for advertising spillovers and flexible unobserved heterogeneity. Assuming that drug advertising only affects the utility of antidepressants implies that its effect on therapy is negative. Not allowing for correlated preferences leads to the erroneous conclusion that drugs and therapy are complements on average, but

[4] To avoid the estimation problem arising from the presence of observed zero market shares, I adopt a Bayesian procedure similar to the one in Li (2019).

[5] "Bundle" is defined as any single- or multiproduct depression treatment option.

has a relatively small effect on the demand elasticity with respect to advertising. Assuming that the time-varying unobservables are uncorrelated across products also leads to estimates that suggest that the products are complements, and results in much larger demand elasticity with respect to advertising.

The paper contributes to two strands of economic literature: on the effects of prescription drug advertising and on discrete-choice models with complementarity. Using TV market borders as a source of exogenous variation in advertising exposure, Shapiro (2019) estimates that antidepressant ads increase depression treatment initiation and improve labor market outcomes by lowering absenteeism. He finds a positive but insignificant effect of drug ads on the use of therapy. Although seemingly contradictory to my findings, this result may be due to the fact that most counties at TV market borders are rural and have few psychotherapy providers.[6]

A few papers find that DTC ads for one drug can have spillover effects on other drugs, but they do not consider the effects on non-drug treatments. Shapiro (2018) uses the TV market border strategy to establish that antidepressant TV ads have a spillover effect on demand for rival products and strong market expansion effect for the category of antidepressant drugs as a whole. Sinkinson and Starc (2019) use the variation in TV advertising of anticholesterol drugs induced by the U.S. election cycle and a regulatory action to find that branded drug ads steal share from other advertising branded drugs but have a spillover effects on non-advertised drugs and market expansion effect on the product category overall.

Other papers have documented that direct-to-consumer advertising can encourage an initial doctor visit or treatment initiation, which is a possible channel for the spillover effect that I find (Hosken and Wendling, 2013; Jayawardhana, 2013; Iizuka and Jin, 2005). Few papers have addressed the complementarity or substitutability of antidepressants and psychotherapy. Berndt et al. (1997) find that higher copayments for therapy are associated with higher use of antidepressants, suggesting that the two treatment options are substitutes. Butikofer et al. (2019) explore the effects of the 2007 FDA black box warnings on antide-

[6]Ellis et al. (2009) document that rural, low-income counties have the lowest number of mental health professionals per capita.

pressants and find that they led to a decline in use of antidepressants and psychotherapy, which is consistent with complementarity.

I also contribute to the literature on discrete-choice demand models with complementarity. In a study of the welfare effects of the entry of the online version of *The Washington Post*, Gentzkow (2007) estimates a model with complementarity and unobserved correlated preferences and establishes the conditions under which the model is identified. I extend his estimation procedure and build on his identification results by addressing the identification challenge that arises if time-varying shocks are correlated across products. Other papers that employ discrete-choice demand models that allow for complementarity are Ershov et al. (2018), Grzybowski and Verboven (2016), Song et al. (2017), and Wakamori (2015).

The rest of the paper is organized as follows. Section 2 provides background information about depression, its treatment options, and advertising. Section 3 describes the data used in the paper. Section 4 develops the demand model and discusses the identification challenges, while Section 5 explains in detail how I take the model to data. Section 6 interprets the results of the estimation, and Section 7 concludes.

2 Empirical Setting

Depression is a mental health disorder characterized by a variety of symptoms including feelings of deep sadness, loss of interest in activities previously enjoyed, suicidal thoughts, and bodily and cognitive changes that affect everyday functioning. Depending on the severity and duration of the symptoms, it can be diagnosed as major depressive disorder (MDD), persistent depressive disorder (dysthymia), or "other" depressive disorder.[7] Depression affects roughly 10% of the adult U.S. population in any 12-month period.[8] In addition to personal

[7] American Psychiatric Association (2013). The "other" depressive disorder category includes cases that meet some but not all criteria for any disorder in the depression category.

[8] Prevalence estimates vary based on data source and criteria used. Brody et al. (2018) report a 12-month prevalence of MDD of 8.1% for the period 2013-2016, which is in line with the prevalence from 2007 to 2012. National Institute of Mental Health (2017) reports 12-month prevalence of dysthymia of 1.5%. Prevalence rates of other depressive disorders are not available either because of data limitations or because it is often incorporated into a broader definition of depression.

suffering, it exacts a substantial economic toll. Greenberg et al. (2015) find that in 2010 the incremental economic cost of major depression was $210.5 billion, including $99 billion in direct medical costs, $9.5 billion in suicide costs, and $102 billion in workplace costs.

There are two main treatment options for depression: pharmacotherapy and psychotherapy. Pharmacotherapy involves the use of prescription drugs, called antidepressants, that target chemicals in the brain that control mood and stress. While antidepressants can improve the symptoms of depression, they often have undesirable side effects. Psychotherapy, or talk therapy, encompasses a variety treatment techniques that can be administered by a psychiatrist, psychologist, psychiatric nurse, or a social worker. The most widely used are cognitive-behavioral, interpersonal, and problem-solving therapies. The goal of these treatments is to help patients change negative thinking patterns, identify factors in their lives that contribute to their depression, and react better to stress.

The medical literature provides evidence that both antidepressants and psychotherapy can be effective in treating depression (Cuijpers et al., 2009; DeRubeis et al., 2008; Friedman et al., 2006). For mild to moderate depression, either treatment option can achieve the desired results (Croghan et al., 1998). As the severity of the disorder increases, treatment guidelines encourage the use of both antidepressants and therapy (Silverman et al., 2015; Crismon et al., 1999). There is some evidence that drugs and therapy work better together. However, the evidence is not sufficient to conclude that they are complements in the health production function, in the sense that using one increases the incremental benefit of using the other.

While both antidepressants and psychotherapy are effective treatments for depression, only antidepressants are advertised on TV. Since 1997, when regulations regarding prescription drug advertising on television were relaxed, pharmaceutical companies have increased their TV ad expenditure significantly. Between 1998 and 2009 ad expenditure increased from $1.2 to $4.5 billion. This type of advertising directly to patients has attracted a lot of controversy. Arguments in favor of DTCA claim that it informs patients about existing treatment options, encourages them to contact a health care provider, reduces stigma associated with

certain conditions, improves adherence to prescribed treatment, and encourages competition between pharmaceutical companies. The arguments against are that ads misinform or present a partial picture of the benefits and risks of the advertised drug, overemphasize drug treatment at the expense of alternative options, encourage inappropriate prescribing, and increase drug prices. As a result, policy makers have proposed a variety of actions toward prescription drug DTCA including an outright ban, removing its preferential treatment in corporate taxes, or stronger regulation.

In the context of depression treatment, psychologists have observed that the share of people diagnosed with depression who use psychotherapy has declined at the expense of antidepressants. Nordbal (2010) notes that between 1997 and 2008 30 percent fewer patients received psychotherapy. While there could be multiple reasons for the decline, including introduction of antidepressants with fewer side effects, changing preferences and treatment styles, increased share of health maintenance organizations that put limits on the amount of therapy patients can receive, one of the alleged reasons for the decline is antidepressant DTC advertising. The purpose of this paper is to shed some light on these claims.

3 Data

3.1 TV Advertising Data

Advertising data come from Kantar Media's Ad$pender database. This data source keeps track of advertising expenditures for more than three million products through multiple marketing media, including television, radio, newspapers, magazines, outdoor, and online. For TV advertising, it provides monthly-level dollar expenditure and number of 30-second advertising segments (or slots). The data are available both at the national (network and cable TV) and local (spot TV) level. Local TV advertising is available for the 101 largest designated market areas (DMAs), the geographical definitions of TV markets.

The total ad expenditure on the class of antidepressant drugs for the 2008–2010 period

is $655 million.[9] In stark contrast to antidepressants, psychotherapy is not advertised on TV.[10]

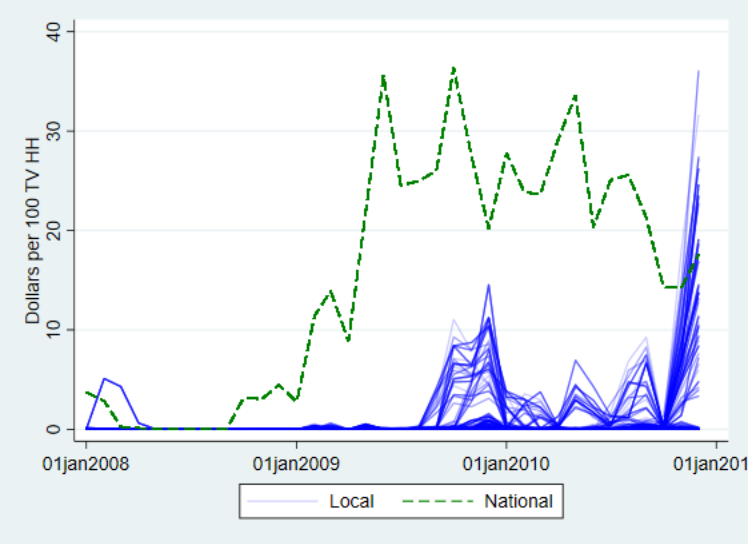

Figure 1: Antidepressant TV Advertising Intensity, National and Local

Notes: Advertising expenditures per 100 TV households for antidepressant prescription drugs at the national level (dashed line) and at the local level for each of the 270 MSAs (solid lines).

To combine national and local ads into a single measurement of advertising intensity, I calculate the ad expenditure per 100 households with a TV set.[11] Total ad intensity in a DMA is the sum of national ad expenditure divided by the number of TV households nationwide (in hundreds) and local ad expenditure divided by the number of TV households in the local market. This is the variable I use for advertising throughout this paper. Figure 1 shows national and local advertising for the period of interest. There is very little advertising activity in 2008, but it picks up in 2009 and 2010. Although national ad intensity is higher

[9]The biggest spenders are Cymbalta, Pristiq, Effexor, Abilify, and Seroquel. Abilify and Seroquel are antipsychotics that were later approved for the treatment of depression as well. I include their ad expenditure in the analysis whenever it is categorized as antidepressant advertising in the Ad$pender database.

[10]The closest category is "Mental Health & Chemical Dependency Clinics," which is predominantly composed of substance abuse clinics and rehabilitation centers. While some of them may provide treatment for depression, including psychotherapy, the focus of the ads is most likely on substance abuse. Furthermore, the total ad expenditure for this category is only 3% of the ad expenditure on antidepressants.

[11]Number of TV households is the unit that ratings company Nielsen often uses in its market size calculations. As of 2010, the United States had a population of 309 million and 115 million TV households for a 2.7 conversion factor from dollars per capita to dollars per TV household.

for most of the time period, there is substantial variation at the local level, which makes the estimation of advertising effects possible.

3.2 Medical and Prescription Drug Claims Data

Data on patient demographics, diagnoses, medical and prescription drug claims, and prices come from the Marketscan Commercial Claims and Encounters Database, which tracks individuals enrolled in employer-provided insurance plans from a convenience sample of large companies. The data cover 395 metropolitan statistical areas (MSAs) for the years 2008 through 2010. Because I am interested in the effect of advertising, I use only the 270 MSAs that can be matched to the 101 DMAs in the Ad$pender dataset.[12]

From the available data, I select individuals who were continuously enrolled for the full three-year period and had complete insurance plan type and demographic information. My sample includes only covered employees because employment information for spouses and dependents is unavailable.[13] I further subset the sample to individuals 18 years of age or older as of January 2008 who were not pregnant during the time period because depression treatment patterns for these subgroups are different.

Table 1 presents a summary of the demographic information for the 2,565,016 individuals that are in the final sample. There is a mixture of salaried and hourly, union and non-union employees, most of whom are full-time. 44.5% are female. The average age is 44.6 years. Preferred provider organization (PPO) is the most widely used type of health insurance plan, followed by health maintenance organization (HMO), point-of-service (POS), and consumer-driven and high deductible health plans (CDHP/HDHP). The sample is geographically representative at the region level.

[12]DMAs are typically centered at large MSA but may include nearby smaller MSAs. For example, the Dallas DMA includes the Dallas-Plano-Irving and Fort Worth-Arlington MSAs.

[13]Depression prevalence and treatment is different among the employed and unemployed. By limiting the analysis to employed individuals I avoid introducing bias from mixing in potentially unemployed individuals at the expense of the generalizability of the findings.

Table 1: Demographics

	%		%
Female	44.5	Northeast	15.96
		Midwest	21.05
Age Group 18-24	3.03	South	37.63
Age Group 25-34	16.76	West	25.36
Age Group 35-44	26.27		
Age Group 45-54	32.74	Hourly	27.60
Age Group 55-64	21.20	Salaried	32.16
HMO	27.40	Union	50.49
POS	15.51	Non-Union	18.72
PPO	50.82		
CDHP/HDHP	6.28	Full-Time	87.01

Note: Based on 2,565,016 covered individuals. HMO = "Health Maintenance Organization", POS = "Point of service", PPO = "Preferred Provider Organization", CDHP = "Consumer Driven Health Plan", HDHP = "High Deductible Health Plan". Employees can be salaried, hourly, or uknown; union, non-union, or unknown; full-time, early retiree, part-time, or other.

Even though antidepressants and psychotherapy can treat a variety of conditions, the focus of this paper is on their use in treating depression. For this reason, I identify individuals diagnosed with three types of depression—major depressive disorder (MDD), dysthymia, and "other" depression—and analyze their choice of depression treatment.[14] Table 2 summarizes the share of the sample that diagnosed with each disorder from 2008 to 2010. Overall, between 4.5% and 5.0% were diagnosed with depression in any given year. This is not inconsistent with an overall depression prevalence of 10% because around one half of depression cases go undiagnosed.

[14]Based on the *International Classification of Diseases, Ninth Edition, Clinical Modifications* classification system, I use codes 296.2 and 296.3 for MDD, code 300.4 for dysthymia, codes 309.0, 309.1 (brief and prolonged depressive reaction), and 311 (depression not elsewhere classified) as other depression. I exclude individuals with comorbid schizophrenia, psychotic depression, or bipolar disorder because both the symptoms of and treatments for these conditions differ substantially from those for depression.

Table 2: Depression Diagnosis Percentages by Year

	2008	2009	2010
Major Depressive Disorder	2.04	2.20	2.29
Dysthymia	0.73	0.77	0.79
Depression-Other	1.70	1.88	1.97
Total (any depression)	4.46	4.85	5.06

Note: Based on 2,565,016 covered individuals. All numbers are percentages. Depression-Other includes depression not elsewhere classified and brief and prolonged depressive reaction.

I assume that each individual makes treatment decisions monthly. For each month, I record whether the individual chose an antidepressant, therapy, both, or neither. Because I am examining the effect of advertising at the category level, I aggregate individual antidepressants into two broad categories: branded and generic.[15] If a 90-day supply of a drug was chosen, I assume that it was consumed from the month of purchase until the month the supply was exhausted. For psychoherapy, I combine individual, family, and group sessions into one good.

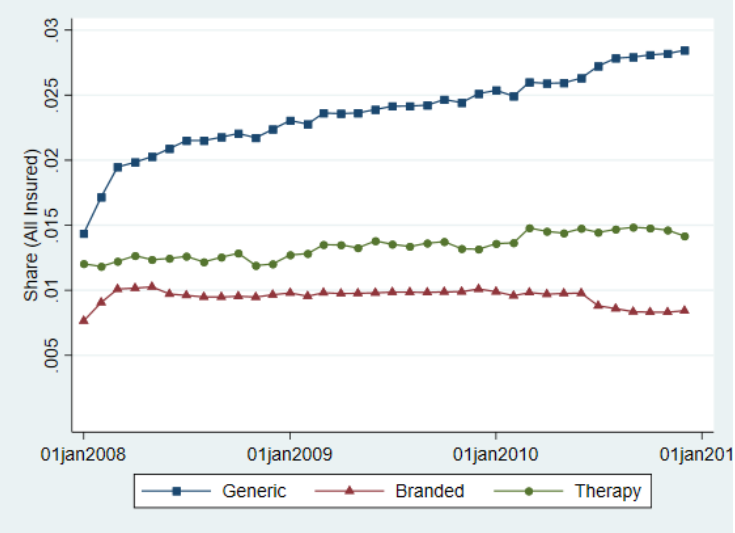

Figure 2: Depression Treatment Product Shares out of All Insured

[15]Only branded antidepressants advertise on TV. Once a generic version of a drug becomes available, the manufacturer of the branded version almost always stops advertising because its market share drops sharply. Ellison and Ellison (2011) find evidence that firms strategically decrease advertising before patent expiration to delay generic entry.

Figure 2 shows the share of all individuals that choose generics, branded drug, and therapy from January 2008 to December 2010. The use of generics increases steadily over the period; at a slower, pace the use of therapy does too. The share of branded drugs is relatively stable for most of the time period until it falls in the second half of 2010.[16] There seems to be a sharp increase in the use of both branded and generic antidepressants in the first three months of the sample. However, this is likely because of 90-day drug supplies purchased in October through December 2007, prior to the beginning of the available data. To avoid understating demand for antidepressants from January to March 2008, I exclude these months from further analysis.

Table 3: Average OOP Prices for Antidepressants and Psychotherapy by Insurance Type

	2008		2009		2010	
	Mean	SD	Mean	SD	Mean	SD
Health Maintenance Organization Plan						
Generic Antidepressants	6.98	2.72	6.61	2.91	6.09	2.50
Branded Antidepressants	26.81	9.76	28.19	9.12	30.13	12.68
Psychotherapy	19.70	10.26	19.88	12.13	22.49	11.88
Point of Service Plan						
Generic Antidepressants	6.85	2.05	6.98	2.51	6.61	2.20
Branded Antidepressants	24.60	7.84	24.82	8.47	25.16	12.16
Psychotherapy	31.56	18.75	34.01	22.14	33.76	17.62
Preferred Provider Organization Plan						
Generic Antidepressants	7.09	1.81	7.01	1.76	6.39	1.50
Branded Antidepressants	26.43	6.49	26.97	6.34	28.60	7.28
Psychotherapy	34.56	18.17	34.15	16.12	34.13	11.95
Consumer Driven/High Deductible Health Plan						
Generic Antidepressants	11.26	11.34	11.10	9.97	11.33	8.65
Branded Antidepressants	46.44	35.64	47.25	35.70	51.93	36.59
Psychotherapy	39.53	29.22	38.90	29.87	42.07	27.46

Note: Prices are in USD, per 30-day supply for drugs and per session for psychotherapy.

Table 3 summarizes the out-of-pocket (OOP) price variation for branded and generic antidepressants and psychotherapy. The OOP price is the dollar amount that the insured patient pays. It may be the full price, if a deductible applies and has not been reached, or just a copay. Because the focus of the paper is on the effect of advertising at the product

[16]This is when generic Effexor XR enters the market.

category level, I calculate a price for each product category by averaging the prices of the component products at the month-state-insurance type level. Table 3 shows that the cheapest treatment option is generic drugs, at around \$6-7 per 30-day supply, while branded drugs and psychotherapy are considerably more expensive, at around \$25 per 30-day supply and \$30 per session, respectively. There is considerable variation around these averages based on insurance plan type, market, and time.

Table 4: Observed and Bayesian Posterior Mean Market Shares

	Mean	SD	Min	10^{th} Pctl	Median	90^{th} Pctl	Max	Share Zeros
# Individuals	9589.9	16922.4	130	482	3201	27082.5	154258	0
Observed Market Shares								
Outside Option	0.9637	0.0121	0.9226	0.9481	0.9648	0.9780	1	0
Generic Only	0.0196	0.0087	0	0.0095	0.0185	0.0306	0.0622	0.003
Branded Only	0.0083	0.0036	0	0.0044	0.0079	0.0129	0.0308	0.014
Therapy Only	0.0054	0.0037	0	0.0014	0.0048	0.0102	0.0379	0.054
Generic & Therapy	0.0033	0.0026	0	0.0004	0.0028	0.0065	0.0197	0.097
Branded & Therapy	0.0015	0.0011	0	0	0.0014	0.0029	0.0090	0.179
Bayesian Posterior Mean Shares								
Outside Option	0.9585	0.0133	0.9066	0.9404	0.9605	0.9738	0.9921	0
Generic Only	0.0202	0.0086	0.002	0.0101	0.0191	0.0314	0.0621	0
Branded Only	0.0090	0.0038	0.0010	0.0049	0.0084	0.0140	0.0368	0
Therapy Only	0.0061	0.0036	0.0004	0.0024	0.0054	0.0108	0.0388	0
Generic & Therapy	0.0040	0.0025	0.0003	0.0017	0.0034	0.0071	0.0203	0
Branded & Therapy	0.0022	0.0013	0.0002	0.0009	0.0019	0.0037	0.0145	0

Notes: Summary statistics based on monthly data for April 2008–December 2010 for 270 MSAs. "# Individuals" is the number of people covered by the Marketscan data in an MSA. The Bayesian posterior mean shares are calculated using the Dirichlet-Multinomial model described in Section 5.6.

Table 4 provides a summary of the shares of each possible treatment choice at the MSA-month level. The positive shares of combination treatment with generic or branded drugs and therapy demonstrate that depressed patients can choose more than one treatment option in a given month. This cannot be the outcome of a typical discrete-choice model in which the consumer selects only the product that provides the highest utility. The model needs to be modified to accommodate choosing multiple alternatives. Furthermore, it is not possible

to determine if the products are complements or substitutes just by looking at the market shares because of the confounding effect of unobserved correlated preferences. To determine the degree of complementarity between products and the effect of advertising on demand, I propose a discrete-choice model that allows for complementarity, spillover effects, and unobserved preference heterogeneity.

4 Demand Model

Antidepressant advertising can have a positive effect on demand for therapy for two reasons. It may boost demand for drugs, which will in turn increase demand for therapy if the two are complements. Alternatively, it may have a spillover effect on therapy—by encouraging people to go to the doctor, for example—and raise demand for it even if the products are substitutes.

To determine the impact of advertising through each channel, I frame the patient's choice of depression treatment as a discrete-choice demand model. For simplicity, I assume that there are three treatment options: outside good (no treatment), drugs, and therapy, $j \in \{0, D, T\}$. Even though in reality the decision is made by a patient and a physician, I assume that their incentives are perfectly aligned and they act as a single decision maker.[17] The individual can choose no treatment, either drugs or therapy alone, or drugs and therapy in combination, $c \in \{0, D, T, DT\}$. The demand system is defined by the following indirect utility functions:

$$\begin{aligned} u_{i0t} &= \varepsilon_{i0t} \\ u_{iDt} &= \delta_D + \beta_D A_{Dt} + \nu_{iD} + \psi_{it} + \varepsilon_{iDt} \\ u_{iTt} &= \delta_T + \beta_T A_{Dt} + \nu_{iT} + \psi_{it} + \varepsilon_{iTt} \\ u_{iDTt} &= (\delta_D + \delta_T + \Gamma) + (\beta_D + \beta_T) A_{Dt} + (\nu_{iD} + \nu_{iT}) + 2\psi_{it} + \varepsilon_{iDTt} \end{aligned} \tag{1}$$

All indirect utilities are normalized by the utility of the outside option. The mean utilities, δ_D

[17] I will call this decision maker an individual or patient for the rest of the paper.

and δ_T, capture the average desirability of drugs and therapy taken alone. The model allows for time-invariant preferences, ν_{iD} and ν_{iT}, time-varying health shocks, ψ_{it}, and idiosyncratic errors, ε_{ict}. Drug advertising, A_{Dt}, enters the utility of both drugs and therapy.[18] A positive coefficient on advertising in the utility of therapy indicates a spillover effect (in utility); a negative coefficient—business stealing. Spillovers can arise if drug ads encourage patients to see their doctor and they decide to take therapy, possibly in combination with drugs, to treat their depression. A business-stealing effect is possible if advertising convinces patients that antidepressants are all they need and discourages them from taking therapy.[19]

The complementarity parameter, Γ, captures the extent to which drugs and therapy work better in combination than on their own. More precisely, Γ is the amount by which the added utility of taking one treatment option changes when the other is taken as well, on average over the idiosyncratic errors ε_{ict}:

$$\Gamma = \mathbb{E}_{\varepsilon}[(u_{iDTt} - u_{iDt}) - (u_{iTt} - u_{i0t})] \tag{2}$$

Complementarity can arise if drugs and therapy treat depression better together than separately or if patients perceive them to do so. If this parameter tends to negative infinity, the drug-therapy bundle is never chosen and the model becomes a traditional discrete-choice model with strict substitutes. Gentzkow (2007) proves that in this type of model if $\Gamma > 0$, the products are complements under the usual definition of complementarity, i.e. demand for one increases when the price of the other decreases (or its utility increases).[20]

Many of the factors involved in the choice of depression treatment are unobservable to the econometrician. Time-invariant unobserved preferences are captured by ν_{iD} and ν_{iT}.

[18] The available data do not allow me to distinguish between persuasive and informative effects of advertising, so I remain agnostic as to which one is present here. Ackerberg (2001) and Ackerberg (2003) provide a reduced-form and a structural approach to identifying informative and persuasive advertising.

[19] In a typical discrete-choice model with substitutes only, a business-stealing effect is present even if drug advertising does not affect the indirect utility of therapy. In a model with complementarity, however, if the products are complements, business-stealing is possible only if drug advertising enters the indirect utility of therapy.

[20] Products are substitutes or independent if $\Gamma < 0$ or $\Gamma = 0$, respectively. See Samuelson (1974) for a comprehensive discussion of various definitions of complementarity.

They may reflect patient attitudes towards the two treatment options that have nothing to do with how effective they are together. If individuals either dislike antidepressants but like psychotherapy (the "hippie" type) or like drugs but not therapy (the "pill-lover" type), these preferences will be negatively correlated in the population. If people either like both or neither, then the correlation in unobserved preferences will be positive.

Another unobservable variable is whether someone is depressed. Individuals go in and out of depression over time but do not always get treatment when they are depressed. While it is reasonable to infer that someone is depressed if they are diagnosed with depression and are taking antidepressants or therapy, it is unclear what their mental health state is if they are not diagnosed or getting treatment. This time-varying, potentially serially correlated, health shock unobservable is embodied by ψ_{it}. When it is in its "depressed" state, it lifts demand for all inside goods. Thus, it is correlated across products.

4.1 Complementarity and Correlated Preferences

The fundamental identification problem in this model is that a large drug-therapy market share can be explained by complementarity between the products or by positively correlated unobserved preferences. Gentzkow (2007) proposes two solutions: an excluded variable that shifts the indirect utility of one product but not the other, and panel data.

The intuition behind the excluded variable approach is simple. If a variable that affects the utility of drugs but not therapy, such as the price of drugs, increases exogenously and demand for psychotherapy decreases, then the two products are complements. With Γ identified, the correlation in preferences is pinned down by the observed shares for each possible treatment choice in the market.

Identification through panel data exploits within-patient treatment shares. Assuming that the ε_{ict} errors are iid extreme value type-1 and integrating them out, it is straightforward to show that the probability of choosing the bundle relative to the product of the probabilities

of choosing each product alone or in combination is:

$$\frac{s_{iDTt}(\nu_{\mathbf{i}}, \psi_{it})}{\Big[s_{iDt}(\nu_{\mathbf{i}}, \psi_{it}) + s_{iDTt}(\nu_{\mathbf{i}}, \psi_{it})\Big]\Big[s_{iTt}(\nu_{\mathbf{i}}, \psi_{it}) + s_{iDTt}(\nu_{\mathbf{i}}, \psi_{it})\Big]} \begin{cases} > 1 & \text{if } \Gamma > 0 \\ = 1 & \text{if } \Gamma = 0 \\ < 1 & \text{if } \Gamma < 0 \end{cases} \tag{3}$$

where $s_{ict}(\nu_{\mathbf{i}}, \psi_{it})$ is an individual's probability of choosing c conditional on the parameters of the model and the particular values of the unobservables $\nu_{\mathbf{i}} = (\nu_{iD}, \nu_{iT})$ and ψ_{it}. If products are complements, a patient consumes the drug-therapy bundle relatively more often.[21] Given a long enough panel and time-varying errors that are not correlated across products, it is possible to identify complementarity even with no variation in an excluded variable.

4.2 Complementarity and Health Shocks

The crucial assumption in the panel data identification approach is that the time-varying shocks are independent across products. If they are not and that is ignored, the correlation in the shocks will be mistaken for complementarity. The situation is analogous to ignoring correlated unobserved preferences in a cross-sectional setting, in which case the correlation is also loaded onto complementarity.

Table 5 provides an example of the problem that unobserved health shocks can create. The health shock can take two values, "high" and "low," and the products are independent. If it were possible to observe the value of the health shocks and if the sequence of choices were long enough, I would be able to calculate the individual-specific ratio in (3) and conclude that the products are independent. However, the health shock is unobservable. If I calculate the ratio in (3) based on all observable data for an individual, I will incorrectly conclude that the products are complements.

[21] If D_{it} is the event that patient i chooses drugs alone or in combination with therapy at time t, and if T_{it} is defined similarly for therapy, I can rearrange equation (3) as $Cov(D_{it}, T_{it}|\nu_i, \psi_{it}) = \mathbb{E}_\varepsilon(D_{it} \cap T_{it}) - \mathbb{E}_\varepsilon(D_{it})\mathbb{E}_\varepsilon(T_{it}) = s_{iDT} - (s_{iD} + s_{iDT})(s_{iT} + s_{iDT}) > 0$ if $\Gamma > 0$. Thus, an equivalent way to express the result in (3) is that the within-individual covariance of the events of choosing drugs and therapy is positive.

Table 5: Identification Challenge with Unobserved Time-Varying Health Shocks

ψ_{it}	s_{iDT}	S_{iD}	S_{iT}	$s_{iDT}/(S_{iD}S_{iT})$
Low	0.04	0.20	0.20	1
High	0.16	0.40	0.40	1
Observed ($\frac{1}{2}$Low + $\frac{1}{2}$High)	0.10	0.30	0.30	$10/9 > 1$

Note: s_{iDT}, $S_{iD} = s_{iD} + s_{iDT}$, and $S_{iT} = s_{iT} + s_{iDT}$ are individual-specific shares of observed choices over time, either conditional on a value of the health shock (first two rows) or not (last row). The last row assumes the number of periods in which the health shock, ψ_{it}, is low and high are equal.

To avoid this pitfall, I allow for time-varying unobservable health shocks, ψ_{it}, that are correlated across products.[22] Getting the distributional assumptions for these shocks is important and puts a caveat to using panel data for identification of complementarity. This puts a greater burden on the excluded variables for identification.

4.3 Advertising Spillovers

The model is simplified substantially by the fact that antidepressants are advertised on TV but psychotherapy is not. This means that there is only one potentially endogenous advertising variable, and a single instrumental variable will be sufficient to address this problem. As long as complementarity and unobserved preferences are separately identified using excluded variables and panel data, it is straightforward to use an instrument that induces exogenous variation in advertising to determine the impact of drug ads on demand for drugs, therapy, and the drugs-therapy bundle.

5 Empirical Implementation

I generalize the model to handle more than two treatment options and more than one market, specify covariates, and parameterize the distributions of the error terms. I then use the fully specified model to derive its likelihood function. Following Goolsbee and Petrin (2004),

[22] Gentzkow (2007) includes a news shock, τ_{it}, in the empirical specification of his model. He justifies it as a way to improve the fit of the model but doesn't discuss its implications about identifying complementarity.

estimation proceeds in two stages. The first stage uses maximum simulated likelihood (MSL) to estimate the coefficients on individual-specific variables and the parameters governing the distributions of the unobservables. Bundle-market-time fixed effects are "concentrated out" as in Berry et al. (2004) and Goolsbee and Petrin (2004), which eases the computational burden substantially. The second stage uses a two-stage least squares regression of the recovered fixed effects to estimate the causal effect of advertising.

5.1 Generalized Model

In month t, individual i, who lives in metropolitan statistical area (MSA) m, chooses a depression treatment option from among no treatment, generic antidepressants, branded antidepressants, and therapy, $j \in \{0, G, B, T\}$. The possible choices are all single- and multiproduct bundles except the ones combining branded and generic drugs, $c \in \{0, G, B, T, GT, BT\}$.[23] The base utility from a single product j is:

$$\begin{aligned} \bar{u}_{ijmt} &= \underbrace{\delta_{jm} + \delta_{jt} + \beta_j A_{Bmt} + \xi_{jmt}}_{\delta_{jmt}} \underbrace{-\alpha P_{ijt} + \bar{\mathbf{X}}_{\mathbf{it}}\bar{\theta}_{\mathbf{j}}}_{\mathbf{X_{ijt}}\theta_{\mathbf{j}}} + \psi_{it} + \nu_{ij} \\ &= \delta_{jmt} + \mathbf{X_{ijt}}\theta_{\mathbf{j}} + \nu_{ij} + \psi_{it} \end{aligned} \tag{4}$$

The base utility consists of two parts: market- and individual-specific. Product mean utility may vary by market and time (δ_{jm}, δ_{jt}) to allow for different treatment styles and changing preferences. Branded antidepressants are the only product advertised on TV. Their advertising, A_{Bmt}, can have a direct effect on branded drugs (β_B) and a business-stealing or spillover effect on generics and therapy (β_G and β_T). Market-level demand shocks unobserved to the econometrician, ξ_{jmt}, may affect pharmaceutical companies' decision how much to advertise in a given market and time period, making advertising potentially endogenous.

Among the individual-specific components of base utility, OOP price, P_{ijt}, varies based on the type of insurance plan. Demographics (age, sex, insurance plan type, employment

[23] Taking more than one antidepressant at a time is strongly discouraged because of possible adverse interactions. Typically, prescribers require a two-week "wash-off" period when switching between antidepressants.

type, diagnosis; $\bar{\mathbf{X}}_{\mathbf{it}}$) can affect each product differently ($\bar{\theta}_{\mathbf{j}}$). For example, women may have different preferences for drugs and therapy than men.

Individual decisions are also affected by unobservable factors: correlated preferences (ν_{ij}) and health shocks (ψ_{it}).I assume that the time-invariant unobserved preferences are distributed multivariate normal with zero mean and unrestricted covariance matrix, $\nu_{\mathbf{i}} \sim MVN(\mathbf{0}, \mathbf{\Sigma})$, where $\nu_{\mathbf{i}} = (\nu_{iG}, \nu_{iB}, \nu_{iT})'$ and $\mathbf{\Sigma}$ is a 3-by-3 symmetric positive definite matrix.[24]

The time-varying health shock is a first-order Markov chain that can take two values:

$$\psi_{it} = \begin{cases} -\infty & \text{if healthy (H)} \\ 0 & \text{if depressed (U)} \end{cases}$$

Healthy individuals have no demand for depression treatment and never purchase any. Depressed individuals, on the other hand, choose their treatment based on the indirect utility of all possible choices. This way of modeling is convenient for three reasons: it is parsimonious; it does not require making arbitrary decisions on which individuals to include in the analysis and for what length of time; and it captures the idea that an individual that does not consume depression treatment can be either depressed or healthy.

The first-order assumption means that the distribution of ψ_{it} depends only on its value in the previous period.[25] Thus, the dynamics of ψ_{it} can be described by a 2-by-2 row-stochastic matrix:[26]

$$\begin{array}{c} \\ H_{it-1} \\ U_{it-1} \end{array} \begin{array}{c} \begin{array}{cc} H_{it} & U_{it} \end{array} \\ \begin{bmatrix} \pi_{HH} & 1-\pi_{HH} \\ 1-\pi_{UU} & \pi_{UU} \end{bmatrix} \end{array}$$

[24] All unobserved preferences, ν_{ij}, are relative to the preference for the outside option. All elements of Σ are normalized by variance and covariance terms of the outside option.

[25] This simplifying assumption makes the model a lot more tractable and allows me to focus on the estimation of complementarity and advertising effects while still acknowledging the transitory nature of depression. In reality, however, transition probabilities are likely endogenous.

[26] A matrix is row-stochastic if the elements of each row are between 0 and 1 and sum to 1.

The first row of the matrix gives the probability that the health shock is in state H (π_{HH}) or state U ($1 - \pi_{HH}$), given that it was in state H in the previous period. The second row provides the analogous probabilities if the previous state was U.

To complete the specification of the health shock, I need to specify the probabilities that it is in each state in the initial period. For simplicity, I assume that the Markov chain is at its stationary (or long-run) distribution for each individual.[27] Thus, the initial-period distribution is $\pi^H = (1 - \pi_{UU})/(2 - \pi_{UU} - \pi_{HH})$ and $\pi^U = 1 - \pi^H$.

With the distributions of the unobservables specified, the conditional indirect utility function for each bundle is defined as:[28]

$$
\begin{aligned}
u_{i0mt} &= \varepsilon_{i0mt} \\
u_{icmt} &= \bar{u}_{ict} + \varepsilon_{icmt} && \text{for } c \in \{G, B, T\} \\
u_{icmt} &= \sum_{j \in c} \bar{u}_{ijt} + \Gamma_{cmt} + \varepsilon_{icmt} && \text{for } c \in \{GT, BT\} \qquad (5) \\
&= \underbrace{\sum_{j \in c} \delta_{jmt} + \Gamma_{cmt}}_{\delta_{cmt}} + \underbrace{\sum_{j \in c} \mathbf{X_{ijt}} \theta_{\mathbf{j}}}_{\mathbf{X_{ict}} \theta_{\mathbf{c}}} + \underbrace{\sum_{j \in c} \nu_{ij}}_{\nu_{ic}} + \underbrace{\sum_{j \in c} \psi_{it}}_{2\psi_{it}} + \varepsilon_{icmt} \\
&= \delta_{cmt} + \mathbf{X_{ict}} \theta_{\mathbf{c}} + \nu_{ic} + 2\psi_{it} + \varepsilon_{icmt}
\end{aligned}
$$

For each bundle, I include idiosyncratic extreme value type-1 error terms, ε_{icmt}, that are independent across patients, products, markets, and time periods.[29] They help rationalize the observed choices and provide closed-form choice probabilities.

I also allow the degree of complementarity to be different for generic drugs and therapy and branded drugs and therapy. Given the restrictions imposed, the complementarity for

[27] This assumption is one way to deal with the initial conditions problem, but is likely too strong. Fortunately, the simplification in estimation that the Markov chain health shock provides does not depend on it. In future versions of this paper, I plan to relax it by modeling the initial-state probabilities as a function of initial-period observables, as in Heckman (1981), or by using a subsample of patients who can reasonably be assumed to be healthy initially, as in Dickstein (2018).

[28] With a slight abuse of notation, I use c to denote a particular choice (singleton or multi-product bundle) and the set of products that the choice contains.

[29] The scale parameter of the ε_{icmt} terms is set to 1, implying a variance of $\frac{\pi^2}{6} \approx 1.64$, to identify the model. The covariance matrix of the correlated preferences is scaled relative to the variance of ε_{ict}.

each market and time period, averaged over the ε_{icmt}'s, is:

$$\Gamma_{cmt} = \delta_{cmt} - \sum_{j \in c} \delta_{jmt} \qquad \text{for } c \in \{GT, BT\} \tag{6}$$

It is necessary to allow complementarity to vary by market and month to fit the observed market shares. Forcing a single average complementarity is easily rejected by the data.

This version of the model allows complementarity to vary by market and time but not by demographics and unobserved preferences. This restriction can be relaxed by allowing the parameters on the demographic variables (and unobservable preferences ν_{ic}) in the multi-product bundle utility functions to be estimated freely. In practice, relaxing it increases the likelihood at convergence modestly at the expense of a large increase in the number of parameters to be estimated.

5.2 Deriving the Likelihood Function

With the model fully specified, I can calculate individual-level choice probabilities. Healthy individuals always choose the outside option:[30]

$$s_{ict}^{H} = \begin{cases} 1 & \text{for } c = 0 \\ 0 & \text{for } c \neq 0 \end{cases} \tag{7}$$

For depressed patients, I analytically integrate out the idiosyncratic error term and derive the choice probabilities conditional on the patient's unobserved preferences:

$$s_{ict}^{U}(\nu_{\mathbf{i}}) = \frac{e^{\delta_{cmt} + \mathbf{X_{ict}}\theta_{\mathbf{c}} + \nu_{ic}}}{1 + \sum_k e^{\delta_{kmt} + \mathbf{X_{ikt}}\theta_{\mathbf{k}} + \nu_{ik}}} \tag{8}$$

Unfortunately, the likelihood of an individual's sequence of choices, even conditional on $\nu_{\mathbf{i}}$, is not simply the product of individual choice probabilities because the health shocks ψ_{it} are not independent over time. To overcome this problem, I need to integrate out the full

[30] Superscripts H and U indicate the value of the patient's health shock, healthy ($\psi_{it} = -\infty$) or depressed ($\psi_{it} = 0$).

sequence of ψ_{it}'s for each patient. In theory, this can be done analytically using the initial state probability and transition probabilities of the Markov chain. Let $\mathbf{c_i} = (c_{i1}, \ldots, c_{i\mathcal{T}})'$ be patient i's sequence of observed choices, where $\mathcal{T}$ is the last time period, and $\Theta = (\delta_{\mathbf{mt}}, \theta, \mathbf{\Sigma}, \pi_{HH}, \pi_{UU})$ be the parameters of the model, where $\delta_{\mathbf{mt}}$ is the vector of all δ_{cmt}'s. The likelihood of an individual's sequence of choices, conditional on $\nu_\mathbf{i}$, is:[31]

$$L_i(\Theta, \nu_i) = Pr(\mathbf{c_i}|\Theta, \mathbf{X}_i, \nu_i) = \sum_{\psi_i \in \Psi} \pi_{\psi_{i1}} s_{ic_{i1}1}^{\psi_{i1}} \pi_{\psi_{i1},\psi_{i2}} s_{ic_{i2}2}^{\psi_{i2}} \ldots \pi_{\psi_{i\mathcal{T}-1},\psi_{i\mathcal{T}}} s_{ic_{i\mathcal{T}}\mathcal{T}}^{\psi_{i\mathcal{T}}} \tag{9}$$

where $\pi_{\psi_{i1}}$ is the initial probability of being in state ψ_{i1} for patient i, $\pi_{\psi_{i1},\psi_{i2}}$ is the transition probability from state ψ_{i1} to ψ_{i2}, and $s_{ic_{it}t}^{\psi_{it}}$ is the probability of observed choice c_{it} in state ψ_{it}. The sum is over all possible health shock sequences $\psi_i = (\psi_{i1}, \ldots, \psi_{i\mathcal{T}})'$.

There are 33 months of data, which implies that there are 2^{33}, or about 8 billion, such sequences. Analytically integrating over all of them by brute force is not feasible. Simulation is an option, but given the length of the panel it will still be computationally expensive and introduce simulation error. Instead, I take advantage of a result from the statistical literature on Hidden Markov Models (HMMs).[32] Using a recursive relationship, the analytical expression for an individual's likelihood function becomes much simpler.

Proposition. *Let f and g stand for any of the values that the first-order Markov chain ψ_{it} can take. Define the joint probability of the observed sequence up to time t and the Markov chain being in state f at time t, conditional on all observables $\mathbf{X_i}$, the unobservable $\nu_\mathbf{i}$, and the parameters of the model Θ:*

$$\phi_{it}(f) = Pr(c_{i1}, \ldots, c_{it}, \psi_{it} = f|\Theta, \mathbf{X}_i, \nu_i)$$

[31] To simplify the notation, $s_{ic_{it}t}^{\psi_{it}}$ stands for $s_{ic_{it}t}^{\psi_{it}}(\nu_\mathbf{i})$.

[32] The name comes from the fact that the unobservable in this type of models is a Markov chain. The seminal paper in the literature on HMMs is Baum and Petrie (1966). A popular tutorial is Rabiner (1989). HMMs have been used heavily in speech recognition and genomic sequencing. Bartolucci et al. (2014) discuss uses of HMMs in economics.

Claim: $\phi_{it}(f)$ *can be computed recursively as:*

$$\begin{aligned}
\phi_{i1}(f) &= \pi_f s^f_{ic_{i1}1} && \text{for } f \in \{H, U\} \\
\phi_{it}(f) &= \Big[\sum_{g \in \{H,U\}} \phi_{it-1}(g)\pi_{gf} \Big] s^f_{ic_{it}t} && \text{for } f \in \{H, U\} \text{ and } t \in \{2, \ldots, \mathcal{T}\}
\end{aligned}$$

Proof. See Appendix A.1. □

Corollary. *The likelihood of an individual's sequence of choices can be calculated as:*

$$\begin{aligned}
L_i(\Theta, \nu_i) &= Pr(c_{i1}, \ldots, c_{i\mathcal{T}} | \Theta, \mathbf{X}_i, \nu_i) \\
&= \phi_{i\mathcal{T}}(H) + \phi_{i\mathcal{T}}(U) \qquad (10)
\end{aligned}$$

Proof. The result follows directly from the Proposition and the definition of $\phi_{i\mathcal{T}}(f)$: sum the joint distribution over the different values of $\psi_{i\mathcal{T}}$ to get the marginal distribution, which is the joint distribution of the observed sequence of actions conditional on observables, $\nu_{\mathbf{i}}$, and parameters. □

Deriving the likelihood in this fashion is known as the *forward algorithm* or α*-pass*. The result is not new, although it is typically applied to models in which the probability of the observed outcome is fixed. In my application, the observed outcome is the result of a discrete-choice model. Conditioning on observables, the unobservable preference vector, and additional model parameters, the result still holds.

Given the likelihood conditional on ν_i, it is theoretically straightforward to derive the unconditional likelihood:

$$L_i(\Theta) = \int_{\nu_i} L_i(\Theta, \nu_i) dF(\nu_i)$$

There is no analytical solution of the integral over ν_i. I simulate it by drawing $R = 100$ 3-vectors $\nu^{\mathbf{r}}_{\mathbf{i}}$ distributed multivariate normal with zero mean and covariance matrix $\mathbf{\Sigma}$ for

each patient, calculating the conditional likelihood, and taking the average:[33]

$$SL_i(\Theta) = \frac{1}{R}\sum_{r=1}^{R} L_i(\Theta, \nu_i^r) \tag{11}$$

With the individual likelihoods taken care of, it is straightforward to calculate the log simulated likelihood of the model:

$$LSL(\Theta) = \log\Big(\prod_i SL_i(\Theta)\Big) = \sum_i \log(SL_i(\Theta)) \tag{12}$$

5.3 Estimation Stage 1: Maximum Simulated Likelihood (MSL)

I estimate the parameters of the model by maximum simulated likelihood.[34] Instead of maximizing over the entire parameter space Θ, however, I "concentrate out" δ_{cmt} and maximize over the rest of the parameters, $\tilde{\Theta} = (\theta, \mathbf{\Sigma}, \pi_{HH}, \pi_{UU})$ as in Goolsbee and Petrin (2004).[35]

For a given $\tilde{\Theta}$, I find δ_{cmt}'s such that predicted bundle shares, $s_{cmt}(\tilde{\Theta}, \delta_{\mathbf{mt}})$, match observed shares, s_{cmt}^{obs}, for each market and month. Berry et al. (1995) (BLP) prove that such δ_{cmt}'s exist and are unique. I use Goolsbee and Petrin's nonlinear least squares optimization approach to estimate the δ_{cmt}'s:[36]

$$\delta_{\mathbf{mt}}(\tilde{\Theta}) = \arg\min_{\delta_{\mathbf{mt}}} \sum_c (s_{cmt}(\delta_{\mathbf{mt}}, \tilde{\Theta}) - s_{cmt}^{obs})^2 \tag{13}$$

Given a vector of δ_{cmt}'s, I use a nonlinear optimization routine to find $\tilde{\Theta}$ that maximizes the log simulated likelihood in (12). The process of finding $\tilde{\Theta}$ and δ_{cmt}'s continues iteratively until convergence.

[33] In practice, I take iid draws from a standard normal u_{ij}, for each j, and calculate $\nu_i = \Lambda \mathbf{u}_i$, where Λ is the lower-triangular Cholesky decomposition of Σ such that $\Lambda\Lambda' = \Sigma$. To draw the u_{ij}'s, I follow the modified Latin hypercube procedure proposed by Hess et al. (2006).

[34] As explained in Greene (2012), Chapter 15, MSL estimation is not consistent for a fixed number of simulation draws. For this reason, I experimented with different numbers of draws and found no substantial difference in the estimates.

[35] The same approach, but in GMM estimation, is used in Berry et al. (2004).

[36] With 270 MSAs, 33 months, and 5 non-empty bundles, there are 44,550 δ_{cmt}'s. Goolsbee and Petrin's approach converges much faster than BLP's contraction mapping.

I calculate predicted market-time shares for each bundle by aggregating individual choice probabilities. Like calculating the individual-level likelihood, calculating individual-level choice probabilities involves evaluating a multidimensional integral, which I do by simulation, re-using the random draws already taken. To integrate over the unobserved health shock, I need the probabilities of being in one of two health states. They can be calculated using the initial-state and transition probabilities. The assumption that the Markov chain is in its steady state, however, implies that the probabilities of being in each state in each period are governed by the stationary distribution of the Markov chain: π^H and $\pi^U = 1 - \pi^H$. Let $\mathcal{I}_{mt}$ be the set of patients in MSA m and month t and N_{mt} be the number of these patients. The predicted share of bundle c is:

$$\begin{aligned} s_{cmt} &= \frac{1}{N_{mt}} \sum_{i \in \mathcal{I}_{mt}} s_{ict} \\ &= \frac{1}{N_{mt}} \sum_{i \in \mathcal{I}_{mt}} \frac{1}{R} \sum_{r=1}^{R} s_{ict}^{r} \\ &= \frac{1}{N_{mt}} \sum_{i \in \mathcal{I}_{mt}} \frac{1}{R} \sum_{r=1}^{R} (\pi^H s_{ict}^{H,r} + (1 - \pi^H) s_{ict}^{U,r}) \end{aligned} \tag{14}$$

Since the healthy type always chooses the outside option, the market shares can be rewritten as:

$$s_{0mt} = \pi^H + (1 - \pi^H) \frac{1}{RN_{mt}} \sum_{i \in \mathcal{I}_{mt}} \sum_{r=1}^{R} s_{i0t}^{U,r} \quad \text{, for } c = 0 \tag{15}$$

$$s_{cmt} = \quad (1 - \pi^H) \frac{1}{RN_{mt}} \sum_{i \in \mathcal{I}_{mt}} \sum_{r=1}^{R} s_{ict}^{U,r} \quad \text{, for } c \neq 0 \tag{16}$$

OOP price varies at the individual level and estimating its effect is part of the first stage. Typically, there is concern that price is endogenous because firms set it based partially on demand factors that are unobservable to the econometrician. This concern is attenuated here because the model includes bundle-MSA-month fixed effects. Any product-level unobservables that vary by MSA-month, which are the typical source of endogeneity, are captured

by the fixed effects. The variation that is left in the OOP is at the insurance plan type. If individuals choose their insurance plan based on their overall demand for health care and if demand for depression treatment is not correlated with that, OOP prices will be exogenous, conditional on the fixed effects.[37]

5.4 Estimation Stage 2: Two-Stage Least Squares

Since I observe advertising at the market-month level, its effect is not separately identified from the bundle-market-month fixed effects δ_{cmt}. However, following Berry et al. (2004), I project the estimated δ_{cmt}'s on advertising, time, and market fixed effects to estimate the own and any possible spillover or business-stealing effects of branded drug advertising. Using the definitions of δ_{cmt} from equations 4 and 5, I set up the regression:

$$\delta_{cmt} = \delta_{cm} + \delta_{ct} + \beta_c A_{Bmt} + \xi_{cmt} \qquad \text{for } c \in \{G, B, T, GT, BT\} \tag{17}$$

Like OOP prices, advertising is set strategically by firms and may be correlated with the error term of the model. To address concerns about endogeneity, I construct an instrument for advertising by calculating the average price for a 30-second TV ad slot in every DMA and month using all TV advertising data in the Ad$pender database.[38] The proposed instrument is relevant because firms have a downward sloping demand for ads.[39] It is also excluded as individual choice of depression treatment is unlikely to be influenced directly by the price of TV ads. Furthermore, given that antidepressant ads are a small portion of all TV advertising, depression treatment demand shocks are unlikely to affect the average price of an ad slot. It is possible, however, that overall viewership may affect the price of a 30-second ad slot

[37] If patients with unobservably high demand for depression treatment select insurance plans with lower OOP prices, the estimated coefficient on price will be biased in a negative direction. To address this issue, I plan to re-estimate the model using individuals on a particular plan type (HMO, PPO, etc.) and use variation in OOPs at the particular plan level (for patients for whom such information is available). If different plans of a given type are sufficiently similar, this approach will eliminate the selection bias.

[38] Such an instrument has been used in Murry (2017) in the study of vertical relationships between car manufacturers and dealers.

[39] Whether an increase in the price of an ad will increase or decrease a firm's advertising expenditure depends on the elasticity of advertising demand with respect to price of advertising. This effect on expenditure will be revealed in the first stage regression of advertising intensity on the price of a 30-second slot.

and demand for depression treatment by making a single TV ad more effective by reaching a wider audience. To deal with this threat to identification, I adjust the average ad price by dividing it by the Nielsen Television Index (NTI) for national broadcast network television programs.[40] This eliminates the variation in prices due to the seasonality in ratings and leaves variation driven by competition for the limited number of ad slots and idiosyncratic factors unrelated to demand for depression treatment. Conditional on the MSA fixed effects and the year and trend time controls I include in the model, I claim that the adjusted ad price is exogenous to demand for depression treatment.[41]

5.5 Identification

As the model is quite complex, it is useful to discuss which moments of the data identify its parameters given the distributional and functional-form assumptions made.

The Markov transition probabilities are pinned down by the observed probabilities of switching from no treatment to consuming any depression treatment and vice versa. The fact that individuals spend long periods with no treatment (because they are healthy) and long periods under treatment (because depression tends to persist) implies that the probabilities of remaining in each state are high and the probabilities of switching are low.

The assumption that healthy individuals do not purchase any depression treatment provides a lot of identifying power because it implies that anyone taking treatment must be unhealthy. Thus, an individual's health state is uncertain only in periods with no depression treatment. For those periods, the Markov assumption helps put a probability on being depressed.[42]

[40]The NTI index is provided by the Television Bureau of Advertising, Inc., at `www.tvb.org`. Because it varies monthly but not across markets, I have to assume that overall TV viewership moves similarly in different markets across the country. Ideally, I would use a market-specific ratings index.

[41]Shapiro (2018) and Sinkinson and Starc (2019) propose alternative identification strategies. I cannot use Shapiro's DMA border identification strategy because my demand data is at the MSA, rather than county, level. While using political advertising is possible, the fact that there is hardly any antidepressant advertising in 2008, when the bulk of the political advertising occurs, means that political ads will be a weak instrument at best.

[42]Hidden Markov Models are particularly useful in speech recognition exactly because of the convenience with which the probability of the unobserved state can be calculated. This "decoding" step is not necessary in the estimation of my model but is useful in thinking about identification.

Complementarity is identified by price variation and the within-individual share of the drug-therapy bundle. If demand for drugs moves in the same direction as demand for therapy in response to a price change, this implies that the two are complements. The same conclusion can be drawn if individuals purchase the bundle relatively more frequently than either product alone.

The coefficients on demographics and prices are identified by the co-variation of these variables and individual-level choices. If women are more likely to take antidepressants but equally likely to go to psychotherapy, then the coefficient on the drugs-female interaction will be positive while the one on therapy-female interaction will be zero.

The parameters governing transitions, complementarity, and demographic and price effects imply certain predicted shares for each bundle in each MSA-month. The extent to which observed shares deviate from the predicted ones identifies the correlation in unobserved patient preferences. For example, if within-patient shares imply that the products are substitutes, but the observed share of the drug-therapy bundle is larger than what the model predicts, this suggests that preferences are positively correlated.

The variance of the unobserved preferences (relative to the normalized variance of the idiosyncratic error term), is identified by the dispersion of purchasing patterns across patients. The presence of patients who always choose drugs but not therapy and others who always choose therapy but not drugs suggests that variances are large. If everyone followed the same treatment plan, the variances would be close to zero.

Finally, the effects of branded antidepressant TV advertising are identified by the covariance between observed choices and the exogenous variation induced by the cost instrument. If advertising leads to greater probability of purchasing drugs, alone or in combination, but not of therapy alone, this implies that there is a positive effect of advertising on drugs but not on therapy.

Functional form and distributional assumption facilitate the estimation of the model, but are not crucial for identification. Berry et al. (2013) show that a nonparametric nonseparable demand system is invertible under the "connected substitutes" condition. This condition

requires that goods are weak gross substitutes and that there are no groups of goods that substitute only among themselves but not to the outside good. Because a discrete-choice demand model with complementarity can be framed as a regular discrete-choice model in which the goods are all possible combinations of the available products, the condition is satisfied. Given that demand is invertible, it is identified in the presence of good instruments. Berry and Haile (2014) provide conditions under which demand is identified with market-level data, while Berry and Haile (2010) focus on situations in which individual-level data are available.[43]

5.6 Zero Observed Market Shares

Table 4 shows that the number of people covered by the Marketscan data varies substantially, with as few as 130 in some MSAs. A prevalence of 10% implies that in such markets there are about 13 depressed individuals. This makes it likely that some of the treatment options will not be chosen in some months. Indeed, for the various bundles this happens in 0.3%–17.9% of all MSA-months. Overall, 21.8% of all MSA-months have at least one zero share.

Estimating the model requires that there are no bundles with zero observed share in any market and time period.[44] Rather than drop MSA-months with zero shares and potentially introduce selection bias, I use a Bayesian procedure similar to the one proposed by Li (2019). Abstracting away from the complexity of the individual-level discrete-choice model, I assume that observed bundle purchase counts in each market and time period, K_{cmt}, are the outcome of a multinomial random variable parameterized by the number of individuals, N_{mt}, and the probabilities for each bundle, p_{cmt} such that $\sum_c p_{cmt} = 1$. Instead of using the observed market shares, which is equivalent to using maximum likelihood to estimate the true underlying probabilities, I use the posterior means from a Bayesian model. I put a weak

[43] Berry and Haile (2016) provide an accessible overview of the identification arguments made in the three papers cited above.

[44] In this model, a zero share implies that for this bundle-MSA-month δ_{cmt} approaches negative infinity, which throws off the estimation procedure. In general, models that involve inverting market shares, such Berry (1994) and Berry et al. (1995), or estimating product-market(-time) fixed effects, such as Goolsbee and Petrin (2004) and Berry et al. (2004), require that there are no zero observed market shares.

and uninformative Dirichlet prior on $\mathbf{p_{mt}}$, which defines the Dirichlet-Multinomial model:[45]

$$K_{cmt} \sim Multinomial(N_{mt}, \mathbf{p_{mt}}) \tag{18}$$

$$\mathbf{p_{mt}} \sim Dirichlet(1, 1, 1, 1, 1, 1) \tag{19}$$

Such models are convenient because the Dirichlet distribution is a conjugate prior to the multinomial likelihood, which means that the posterior distribution is also Dirichlet with parameters $1 + K_{cmt}$ for each c. The posterior mean for each bundle is easy to derive:

$$\hat{p}_{cmt} = \frac{1 + K_{cmt}}{\sum_k (1 + K_{kmt})} = \frac{1 + K_{cmt}}{6 + N_{mt}} \tag{20}$$

The posterior mean is strictly positive, which solves the problem of zero shares and makes estimation possible without discarding any markets or time periods.

6 Results

Even with the use of the forward algorithm to avoid simulating the time-varying health shocks, the model is computationally challenging. To ease the burden, I estimate it using a panel spanning 33 months for a random subsample of 13,500 individuals from the 270 MSAs.

6.1 Price, Demographics, Distribution of Unobservables

Table 6 contains the results from the first stage of the model. These parameter estimates reflect the preferences of depressed patients. By definition, healthy individuals do not consume depression treatment.

[45] The Dirichlet distribution is a multivariate generalization of the beta distribution and is suitable as a prior for probability vectors because it is defined on the unit simplex. The prior is weak because the hyperparameters (the vector of 1's) are small in magnitude. It is uninformative because it implies that vector of multinomial probabilities is equally likely. Both of these characteristics of the prior allow the observed data to be the main determinant of the posterior distribution. A similar Dirichlet-Multinomial model has been used in Conlon and Mortimer (2019) in the empirical study of diversion ratios.

Table 6: First Stage (MSL) Estimates

	Est	SE	Est	SE	Est	SE
OOP	-0.206	(0.143)				
×MDD	0.721***	(0.083)				
×Depression-Other	0.039	(0.093)				
×Female	-0.041	(0.101)				
×Age	0.103**	(0.048)				
×Salaried	0.221	(0.14)				
×Union	0.283**	(0.136)				
	×Generic		×Branded		×Therapy	
Female	0.958***	(0.095)	1.106***	(0.142)	1.134***	(0.134)
Age	0.363***	(0.046)	0.132*	(0.08)	-0.202***	(0.068)
Age^2	-0.174***	(0.051)	-0.066	(0.058)	-0.144***	(0.041)
HMO	0.277	(0.181)	-0.254	(0.205)	1.025***	(0.233)
POS	-0.229	(0.198)	-0.151	(0.219)	1.456***	(0.223)
PPO	0.518***	(0.161)	1.232***	(0.189)	1.779***	(0.212)
Salaried	-0.574***	(0.133)	-0.114	(0.206)	-1.008***	(0.206)
Hourly	-1.099***	(0.126)	-0.571***	(0.141)	-0.706***	(0.116)
Union	1.148***	(0.144)	0.971***	(0.204)	0.784***	(0.178)
Non-Union	1.079***	(0.125)	1.29***	(0.13)	0.841***	(0.103)
Full-Time	-0.347***	(0.117)	0.066	(0.137)	-0.668***	(0.101)
MDD	0.487***	(0.066)	0.301**	(0.117)	0.335***	(0.102)
Depression-Other	0.662***	(0.078)	1.722***	(0.149)	-0.589***	(0.135)
Log-likelihood			-31,839.5			
Observations			445,500			

Notes: Bundle-MSA-month fixed effects included. Demographics are interacted with product dummies, OOP and OOP interactions are not. For the multiproduct bundles (not shown), the demographic effects equal the sum of the demographic effects on the component single-product bundles. OOP prices and age are measured in standard deviations ($24.52 and 10.62 years, respectively) and centered around their means ($26.70 and 44.85 years). Employees can be salaried, hourly, or unknown; unionized, non-unionized, or unknown; full-time, early retiree, part-time, or "other." Depression diagnosis can be major depressive disorder (MDD), dysthymia, or depression-other. The omitted health insurance type is CDHP/HDHP. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Taking advantage of the individual-level data, I allow the price sensitivity to vary by demographic characteristics. The base group of individuals seems to dislike higher prices, although the coefficient is not statistically significant. While men and women are equally price sensitive, older individuals and those who are salaried or unionized employees seem less sensitive. This could partially be driven by the fact their younger, part-time or non-

unionized counterparts are likely to have lower income, which I do not observe. Strangely, patients diagnosed with major depressive disorder seem to derive greater utility from higher prices. This result should be taken with a grain of salt. It is likely driven by the fact that depression severity varies unobservably even within major depression and more severely depressed patients choose more expensive treatments.[46]

In addition to the interaction with price, the type of diagnosis affects choice directly. Patients with MDD have higher demand for all types of treatments, while those with other forms of depression are more heavily reliant on antidepressants and less on psychotherapy compared to patients with dysthymia.[47]

Women use depression treatment more heavily, which has been documented in the medical literature. Generic and branded antidepressant use increases at a decreasing rate until around age 56, whereas therapy use is heaviest among 37-year-olds and decreases away from this age. Relative to hourly workers, salaried ones use more antidepressants and slightly less therapy, although both groups use less depression treatment than workers with unknown status. Unionized and non-unionized workers use approximately the same amount of depression treatments, which is more than workers with unknown unionization status. Full-time employees fill fewer generic antidepressant prescriptions and take less psychotherapy compared to (mostly) early retirees.

Finally, PPO plans are associated with the highest level of depression treatment usage, although the results cannot tell if this is due to selection or because PPO plans have more generous coverage. HMO, POS, and CDHP/HDHP plans have approximately the same utilization of generic and branded antidepressants. In terms of psychotherapy use, PPO is the highest, followed by POS, HMO, and CDHP/HDHP plans. This can be explained partially by the ease with which patients can access therapists under each plan—PPO plans do not require a referral from a primary care physician, whereas POS and HMO plans do.

[46]To address this problem, I plan to use more granular diagnosis information, which I am currently aggregating over.

[47]The current version of the model assumes that the degree of complementarity does not depend on the type of diagnosis. This can be relaxed in future versions.

Table 7: Covariance/Correlation Matrix of Unobserved Preferences

	Generic	Branded	Therapy
Generic	8.64	0.34	0.25
Branded	4.06	16.25	0.44
Therapy	2.34	5.51	9.84

Notes: Variance and covariance terms on the main diagonal and below; correlation coefficients—above. All parameter estimates are significant at the 1% significance level.

The estimated covariance matrix of the unobserved preferences for each product, ν_{ij}, is in Table 7. The variance terms (8.6, 16.3, 9.8 for generics, branded, and therapy, respectively) are much larger than the normalized variance of the extreme value type-1 error of 1.64, which suggests that unobservable factors play a much larger role in the choice of depression treatment than observable characteristics. It also means that observably similar patients choose radically different treatment plans: from exclusively pharmacologic to exclusively psychotherapeutic treatment and anywhere in-between. Neither of these results is surprising given how idiosyncratic the manifestation of depression and the effectiveness of different treatments are.

Table 8: Markov Chain Health Shock Transition Probabilities and Stationary Distribution

Panel A: Transition probabilities		
	Healthy_{t+1}	Depressed_{t+1}
Healthy_t	0.9982	0.0018
	(0.0000)	(0.0000)
Depressed_t	0.0175	0.9825
	(0.0000)	(0.0000)
Panel B: Stationary distribution		
	Healthy	Depressed
Long-run share	0.9066	0.0934

Note: Standard errors in parentheses.

The positive covariance terms imply that patients tend to either like any two treatment options or neither. They also suggest that patients see all three treatment options as somewhat "similar" in the sense that a change in the price of one would lead to greater substitution

to inside bundles than to the outside option compared to what an IIA logit model would predict.[48]

The estimated health state transition probabilities, shown in Table 8, indicate that both the healthy and unhealthy states are highly persistent. The probability of falling into depression, having been healthy the previous month, is 0.18% whereas the probability of recovery once in a depression is 1.75%. The implied stationary distribution of healthy and unhealthy people is 90.66% healthy and 9.34% depressed. The share of depressed is higher than what is actually observed, as reported in Table 2, but there is no inconsistency because the former includes both diagnosed and undiagnosed cases whereas the observed share includes only diagnosed cases. Given that roughly half of depression cases go undiagnosed, the results are in line with the medical literature (Williams et al., 2017).

6.2 Complementarity

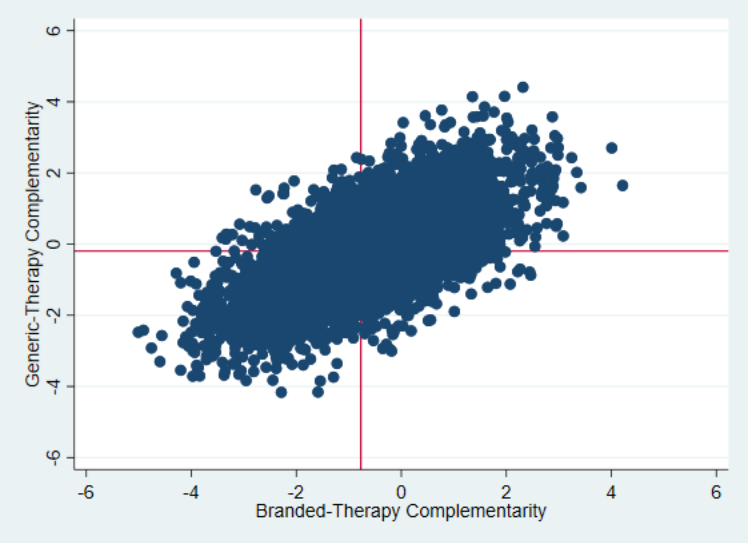

Figure 3: Generic-Therapy and Branded-Therapy Complementarity by MSA and Month

Note: Each point represents a $(\Gamma_{GTmt}, \Gamma_{BTmt})$ pair for a particular MSA-month. The vertical line is at the mean branded-therapy complementarity (-0.77), the horizontal—at the mean generic-therapy complementarity (-0.19).

[48] Greater substitution to inside bundles, however, does not necessarily mean that the products comprising the bundles are substitutes. That is determined by the degree of complementarity.

Equation 6 defines the market- and time-specific complementarity parameters for the generic-therapy and branded-therapy combination treatments. Using the estimated bundle-MSA-month fixed effects, I calculate the implied complementarities and plot them in Figure 3. If there were a single complementarity parameter for each multiproduct bundle, the plot would consist of a single point. The wide dispersion in the plot suggests that the degree of complementarity varies over markets and time. The reason for that could be that patients have different preferences, physicians follow different treatment guidelines, or there are idiosyncratic factors that shift complementarity.

While both generic and branded antidepressants are substitutes to psychotherapy on average, there's substantial variation over product categories, markets, and time. As indicated by the average complementarity parameters, branded drugs are more substitutable with therapy ($\Gamma_{BT} = -0.77$) than generic drugs ($\Gamma_{GT} = -0.19$). Furthermore, Figure 4 reveals the distribution of average (over time periods) complementarity at the MSA level: in 113 out of the 270 MSAs (42%) generic drugs and therapy are complements; for branded drugs, there are 42 such MSAs (16%).

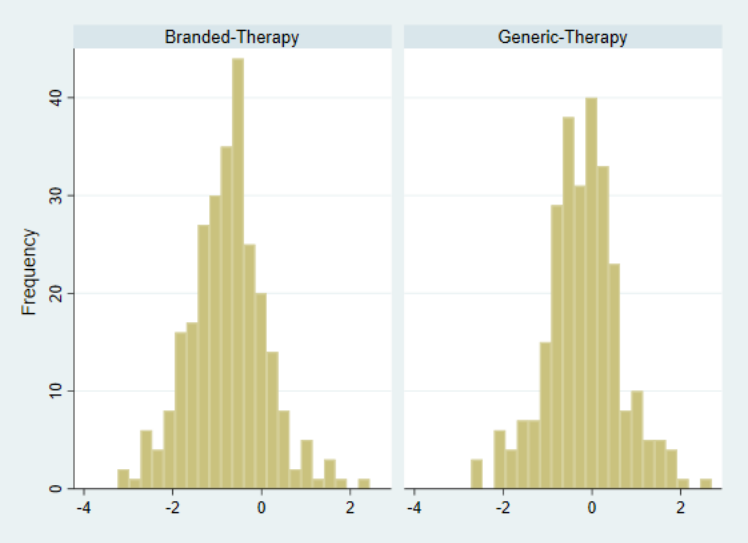

Figure 4: Distribution of Average MSA Complementarity

Figure 5 shows that the average complementarity across MSAs decreased over the 2008–

2010 period, which suggests that patients became less likely to take combination treatment over time. If this trend had been going on for a while, it might partially explain the fall of psychotherapy's share over time.

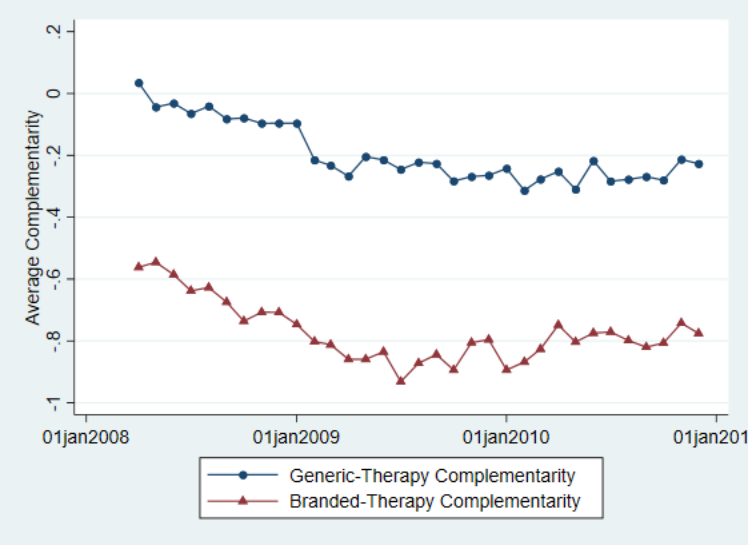

Figure 5: Average Generic-Therapy and Branded-Therapy Complementarity Over Time

6.3 Advertising Effects

The first stage of the estimation recovers the bundle-MSA-month fixed effects, which reveals the average degree of complementarity and its distribution. Advertising effects, however, are not separately identified from the fixed effects. To estimate the effects of advertising, I project the estimated δ_{cmt}'s on advertising intensity, MSA and year fixed effects, and a time trend.

Ordinary least squares (OLS) results are presented in Table 9. An increase in advertising intensity is associated with higher utility for branded drugs, lower utility for generics, and no effect for therapy. Because branded drugs are substitutes with both generics and therapy, this implies that branded drug TV ads increase demand for branded drugs overall and decrease it for generics and therapy.

Table 9: Second Stage, OLS Estimates

	Generic	Branded	Therapy	G-T Bundle	B-T Bundle
δ_c	-2.8545***	-5.8453***	-5.1612***	-8.2096***	-11.7781***
	(0.0673)	(0.0789)	(0.072)	(0.0813)	(0.0889)
Ads ($/100 TV HH)	-0.0026***	0.0059***	0.0009	-0.0017**	0.0068***
	(0.0006)	(0.0007)	(0.0006)	(0.0007)	(0.0008)
Time effects	Year-Trend	Year-Trend	Year-Trend	Year-Trend	Year-Trend
Market FEs	MSA	MSA	MSA	MSA	MSA
Obs	8,910	8,910	8,910	8,910	8,910

Notes: Standard errors in parentheses. δ_c is the mean utility for a bundle for the base time period and MSA. For the multiproduct bundles, the advertising effects equal the sum of the effects on the component single-product bundles. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

To address potential endogeneity concerns, I use the average price of a 30-second TV advertising segment, adjusted for ratings, as an instrument. Panel B of Table 10 shows the outcome of the first stage of the two-stage least squares (2SLS) procedure. The relationship between advertising and the instrument is negative, as expected. The partial F-statistic of 505 is much larger than conventionally used cutoff points, indicating that the instrument is strong.

Table 10: Second Stage, 2SLS Estimates

Panel A: regression of bundle-MSA-month fixed effects on advertising					
	Generic	Branded	Therapy	G-T Bundle	B-T Bundle
δ_c	-2.8148***	-5.8388***	-5.1303***	-8.139***	-11.7407***
	(0.0689)	(0.0803)	(0.0735)	(0.0833)	(0.0906)
Ads ($/100 TV HH)	0.0053***	0.0072***	0.0071***	0.0124***	0.0143***
	(0.0025)	(0.003)	(0.0027)	(0.0031)	(0.0033)
Time effects	Year-Trend	Year-Trend	Year-Trend	Year-Trend	Year-Trend
Market FEs	MSA	MSA	MSA	MSA	MSA
Obs	8,910	8,910	8,910	8,910	8,910
Panel B: first stage of instrumental variables regression					
Adj Avg Price per Ad			-0.0057		
Partial F-stat			504.8		

Notes: Standard errors in parentheses. δ_c is the mean utility for a bundle for the base time period and MSA. For the multiproduct bundles, the advertising effects equal the sum of the effects on the component single-product bundles. The average price per ad is adjusted for TV ratings. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

The second-stage 2SLS results are in Panel A of Table 10.[49] The estimates suggest that advertising makes branded drugs more desirable, as in the OLS version. In the instrumented version of the model, however, there is evidence of advertising spillovers in utility: advertising significantly increases the utility from generic antidepressants and therapy as well. The difference compared to the OLS results suggests that firms may be advertising more heavily in markets with lower demand, especially for generics and therapy.

The temptation to interpret the positive advertising coefficients as an indication that demand for all three products (therapy, generic and branded drugs) increases in response to higher advertising should be resisted. The overall effect of advertising depends both on these coefficients and on the degree of complementarity between the products. Given that generic and branded drugs are substitutes to therapy, if there were no advertising spillovers in utility but only a positive effect on branded drugs, an increase in advertising intensity would increase demand for the branded-only and branded-therapy bundles and decrease demand for the therapy-only, generics-only, and generics-therapy bundles. Despite the positive effect on the branded-therapy bundle, the overall effect on demand for therapy would be negative. The presence of advertising spillovers in utility, however, dampens the substitution away from therapy and, if the spillover is strong enough, may boost demand for therapy overall.

Table 11: Demand Elasticities with Respect to Advertising

Bundles	Elasticity	Products	Elasticity
Outside Option	-0.0053	Outside Option	-0.0053
Generic-Only	0.0033	Generic	0.0050
Branded-Only	0.0063	Branded	0.0076
Therapy-Only	0.0061	Therapy	0.0093
Generic-Therapy	0.0147		
Branded-Therapy	0.0177		

Note: Demand elasticities with respect to branded drug advertising are calculated by averaging individual-level elasticities over patients and time periods.

[49] I have also estimated the model using three-stage least squares (3SLS), which uses the covariance in the errors of the equations for each bundle to enhance estimation efficiency. The results are similar, with slightly smaller standard errors. The downside of 3SLS is that misspecification in one equation is transferred to the entire system of equations. Because of this, I chose the less efficient but more robust estimation method.

Table 11 presents the estimated average (over patients and time periods) elasticities for each bundle and each product. They indicate that advertising lifts all individual inside bundles and, as a result, each product overall. Thus, the advertising spillovers in utility translate into spillovers in demand as well: even though branded and generic antidepressants are substitutable with psychotherapy, the overall effect of advertising is to increase demand for all three of them.

The effect of advertising is positive but modest in magnitude. The estimates imply that a 10% increase in advertising increases demand for generic drugs, branded drugs, and therapy by 0.050%, 0.076%, and 0.093%, respectively.[50] The share of the outside option shrinks by 0.053% in response to the same increase in advertising. The numbers are slightly lower, although largely in line, with other estimates from the literature.[51]

6.4 Importance of Assumptions on the Unobservables

There are three features of the model that are crucial for estimating the correct degree of complementarity and advertising elasticities: allowing for time-invariant correlated preferences, time-varying health shocks, and advertising spillovers. I evaluate their importance by eliminating them from the model one at a time and re-estimating it.

Because advertising is subsumed by the bundle-MSA-month fixed effects, eliminating spillovers does not change the results from the first stage of the estimation. Thus, the effects of price and demographics on utility, the covariance matrix of the unobserved preferences, and the Markov transition probabilities remain the same as in the main specification. It also implies that the average level of complementarity and the market-time deviations from it are the same as in the main model. The differences appear in the effect of advertising on utility and especially in the elasticities as shown in Tables 16 and 17, column (2), in

[50] These are contemporaneous effects. Additional specifications suggest that advertising has no significant lagged effect on branded and generic antidepressants. The advertising effect on psychotherapy, however, persists for an additional month or two.

[51] Sinkinson and Starc (2019) find that the category of statin drugs expands by 0.13% in response to a 10% increase in advertising. Shapiro (2018) finds that the outside share in the market for antidepressants decreases by between 0.08% and 0.23% in response to the same change in advertising. Caution should be used for these last comparisons since the outside option is defined differently in Shapiro's model and mine.

Appendix A.2. The effect of advertising on branded drugs is somewhat higher than in the main model, although within a standard deviation from it. The most significant difference, however, is in the implied elasticities. While the main results suggest that advertising "lifts" all antidepressants and therapy, the model with no spillovers indicates that branded drugs benefit at the expense of generics, therapy, and the outside option. The results are driven by the fact that both generics and therapy are substitutes for branded drugs and even though the branded-therapy bundle benefits from advertising, this effect is not strong enough to offset the decline in therapy-only and generic-therapy. The fact that the ad effects on generics and therapy flip in sign underscores the importance of allowing for spillovers.

Column (3) in Tables 12–17 shows the results for a model that allows for random coefficients on generics, branded drugs, and therapy but restricts them to be uncorrelated, while maintaining the other features of the main model. The results from the MSL estimation are qualitatively similar to those from the main specification. The difference in the log-likelihood at convergence is relatively small, which suggests that allowing for correlated preferences affects the overall fit relatively little. Most significantly, every single Γ_{cmt} is larger than in the baseline model. As a result, the average level of complementarity increases substantially and implies that therapy and both types of antidepressants are complements on average. This, however, translates into relatively modest changes in the estimated average elasticities. In response to a 10% increase in advertising, demand for generics, branded drugs, and therapy increases by 0.051%, 0.089%, and 0.097%, which is respectively 2%, 17%, and 4% higher than the main model.

Finally, column (4) in Tables 12–17 shows the results from the model in which there are no health shocks, so that the only time-varying unobservables are the extreme value type-1 error terms, but spillovers and correlated preferences are allowed. The first stage of the estimation indicates that the correlation in preferences is much larger than in the main specification. This is expected—the ignored positive correlation in the health shocks is loaded (partially) on the correlation of the unobserved preferences. The log-likelihood at convergence indicates a significantly worse fit than the main model. Like ignoring correlation

in the time-invariant unobservable preferences, ignoring time-varying health shocks increases each individual Γ_{cmt}, which once again leads to the conclusion that drugs and therapy are complements on average. In this specification, however, there is also a significant increase in the average elasticities—from 6 to 13 times. The reason for this result is that this version of the model assumes that all individuals in the sample are in the market for depression treatment whereas the main model estimated a sizeable portion of healthy individuals that have no demand for depression treatment. Even though not all individuals are advertising-marginal, the share that are is much larger. This underscores the importance of time-varying product-correlated shocks in discrete-choice models with complementarity that use panel data.

7 Conclusion

I study the effect of antidepressant advertising on demand for depression treatment using a discrete-choice model that allows for complementarity and advertising spillovers. The model allows for flexible unobserved heterogeneity: time-invariant preferences and time-varying health shocks, both of which can introduce correlation in utility across products. To separately identify complementarity from unobserved correlated preferences, I use panel data on choices and variation in an excluded variable, price. I estimate the causal effect of advertising by using a cost-based instrument. The model advances existing discrete-choice models with complementarity by allowing for advertising spillovers, multiple markets, and endogenous variables; discussing the threats to identification arising from time-varying, product-correlated unobservables; and modeling such unobservables in a computationally feasible way.

The results indicate that branded drug TV advertising increases demand for psychotherapy. This is the outcome of two forces working in opposite directions. First, drugs and therapy are substitutes, which implies that advertising, which boosts demand for drugs, should decrease demand for therapy. However, advertising has a spillover effect on the utility of therapy. This spillover effect dominates the substitution effect for a net positive impact

on psychotherapy.

This result has important policy implications. First, providers of psychotherapy, who feel that antidepressant ads are stealing their patients, need not worry—drug advertising actually helps them. Second, policymakers that propose banning or curtailing prescription drug ads need to be aware of the unintended consequences of such actions. Shapiro (2019) shows that antidepressant advertising improves labor market outcomes and that the benefits far outweigh the costs. This paper sheds additional light on one of the channels through which the effect occurs.

A direction for future work is to add a supply-side model and study the effects of a counterfactual ban of prescription drug TV ads. This will require the demand analysis to proceed at the product level, which is possible but will be more computationally cumbersome. It will also require data on firms detailing, or direct-to-physician advertising, which may be a substitute or complement to direct-to-consumer advertising for firms.

Another potentially fruitful application of this type of demand model with complementarity and unobserved heterogeneity is the study of the welfare effects of tying and bundling. The effects of these pricing practices depend both on the degree of complementarity and the correlation and preferences, which can be estimated with the model I propose.

References

Ackerberg, D. A. (2001): "Empirically Distinguishing Informative and Prestige Effects of Advertising," *The RAND Journal of Economics*, 32, 316–333.

——— (2003): "Advertising, Learning, and Consumer Choice in Experience Good Markets: An Empirical Examination," *International Economic Review*, 44, 1007–1040.

American Psychiatric Association (2013): *Diagnostic and Statistical Manual of Mental Disorders*, Arlington, VA: American Psychiatric Association, 5th ed.

Bartolucci, F., A. Farcomeni, and F. Pennoni (2014): "Latent Markov Models: A Review of a General Framework for the Analysis of Longitudinal Data with Covariates," *TEST*, 23, 433–465.

Baum, L. E. and T. Petrie (1966): "Statistical Inference for Probabilistic Functions of Finite State Markov Chains," *Annals of Mathematical Statistics*, 67, 1554–1563.

Berndt, E. R., R. G. Frank, and T. G. McGuire (1997): "Alternative Insurance Arrangements and the Treatment of Depression: What Are the Facts?" *American Journal of Managed Care*, 3, 243–250.

Berry, S., A. Gandhi, and P. Haile (2013): "Connected Substitutes and Invertibility of Demand," *Econometrica*, 81, 2087–2111.

Berry, S. and P. Haile (2010): "Nonparametric Identification of Multinomial Choice Demand Models with Heterogenenous Consumers," *Working paper*.

——— (2016): "Identification in Differentiated Products Markets," *Annual Review of Economics*, 8, 27–52.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841–890.

——— (2004): "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market," *Journal of Political Economy*, 112, 68–105.

Berry, S. T. (1994): "Estimating Discrete-Choice Models of Product Differentiation," *The RAND Journal of Economics*, 25, 242–262.

Berry, S. T. and P. A. Haile (2014): "Identification in Differentiated Products Markets Using Market Level Data," *Econometrica*, 82, 1749–1797.

Brody, D. J., L. A. Pratt, and J. P. Hughes (2018): "Prevalence of Depression Among Adults Aged 20 and Over: United States, 2013–2016," *NCHS Data Brief*.

Butikofer, A., C. Cronin, and M. Skira (2019): "Employment Effects of Healthcare Policy: Evidence from the 2007 FDA Black Box Warning on Antidepressants," *Working paper*.

Conlon, C. and J. H. Mortimer (2019): "Empirical Properties of Diversion Ratios," *Working paper*.

Crismon, M. L., M. Trivedi, T. A. Pigott, A. J. Rush, R. M. A. Hirschfeld, D. A. Kahn, C. DeBattista, J. C. Nelson, A. A. Nierenberg, H. A. Sackeim, and M. E. Thase (1999): "The Texas Medication Algorithm Project: Report of the Texas Consensus Conference Panel on Medication Treatment of Major Depressive Disorder," *The Journal of Clinical Psychiatry*, 60, 142–156.

Croghan, T. W., C. A. Melfi, W. E. Crown, and A. Chawla (1998): "Cost-Effectiveness of Antidepressant Medications," *The Journal of Mental Health Policy and Economics*, 1, 109–117.

Cuijpers, P., A. van Straten, L. Warmerdam, and G. Andersson (2009): "Psychotherapy versus the Combination of Psychotherapy and Pharmacotherapy in the Treatment of Depression: A Meta-Analysis," *Depression and Anxiety*, 26, 279–288.

DeRubeis, R. J., G. J. Siegle, and S. D. Hollon (2008): "Cognitive Therapy versus Medication for Depression: Treatment Outcomes and Neural Mechanisms," *Nature Reviews Neuroscience*, 9, 788–796.

Dickstein, M. J. (2018): "Efficient Provision of Experience Goods: Evidence from Antidepressant Choice," *Working paper*.

Ellis, A. R., T. R. Konrad, K. C. Thomas, and J. P. Morrissey (2009): "County-Level Estimates of Mental Health Professional Supply in the United States," *Psychiatric Services*, 60, 1315–1322.

Ellison, G. and S. F. Ellison (2011): "Strategic Entry Deterrence and the Behavior of Pharmaceutical Incumbents Prior to Patent Expiration," *American Economic Journal: Microeconomics*, 3, 1–36.

Ershov, D., J.-W. P. Laliberté, and S. Orr (2018): "Mergers in a Model with Complementarity," *Working paper*.

Friedman, M. A., J. B. Detweiler-Bedell, H. E. Leventhal, R. Home, G. I. Keitner, and I. W. Miller (2006): "Combined Psychotherapy and Pharmacotherapy for the Treatment of Major Depressive Disorder," *Clinical Psychology: Science and Practice*, 11, 47–68.

Gentzkow, M. (2007): "Valuing New Goods in a Model with Complementarity: Online Newspapers," *The American Economic Review*, 97, 713–744.

Goolsbee, A. and A. Petrin (2004): "The Consumer Gains from Direct Broadcast Satellites and the Competition with Cable TV," *Econometrica*, 72, 351–381.

Greenberg, P. E., A.-A. Fournier, T. Sisitsky, C. T. Pike, and R. C. Kessler (2015): "The Economic Burden of Adults With Major Depressive Disorder in the United States (2005 and 2010)," *The Journal of Clinical Psychiatry*, 76, 155–162.

Greene, W. H. (2012): *Econometric Analysis*, Upper Saddle River, NJ: Prentice Hall, 6th ed.

Grzybowski, L. and F. Verboven (2016): "Substitution between Fixed-Line and Mobile Access: The Role of Complementarities," *Journal of Regulatory Economics*, 49, 113–151.

Heckman, J. J. (1981): "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process," in *Chapter 4 in Structural Analysis of Discrete Data, Ed. by Charles Manski and Daniel McFadden*, Cambridge, MA: MIT Press.

Hess, S., K. E. Train, and J. W. Polak (2006): "On the Use of a Modified Latin Hypercube Sampling (MLHS) Method in the Estimation of a Mixed Logit Model for Vehicle Choice," *Transportation Research Part B: Methodological*, 40, 147–163.

Hosken, D. and B. Wendling (2013): "Informing the Uninformed: How Drug Advertising Affects Check-up Visits," *International Journal of Industrial Organization*, 31, 181–194.

Iizuka, T. and G. Z. Jin (2005): "The Effect of Prescription Drug Advertising on Doctor Visits," *Journal of Economics & Management Strategy*, 14, 701–727.

Jayawardhana, J. (2013): "Direct-to-Consumer Advertising and Consumer Welfare," *International Journal of Industrial Organization*, 31, 164–180.

Li, J. (2019): "Compatibility and Investment in the U.S. Electric Vehicle Market," *Working paper*.

Murry, C. (2017): "Advertising in Vertical Relationships: An Equilibrium Model of the Automobile Industry," *Working paper*.

National Institute of Mental Health (2017): "Persistent Depressive Disorder (Dysthymic Disorder)," https://www.nimh.nih.gov/health/statistics/persistent-depressive-disorder-dysthymic-disorder.shtml.

Nordbal, K. (2010): "Where Has All the Psychotherapy Gone?" *Monitor on Psychology*, 41.

Rabiner, L. R. (1989): "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proceedings of IEEE*, 77, 267–295.

Samuelson, P. A. (1974): "Complementarity: An Essay on The 40th Anniversary of the Hicks-Allen Revolution in Demand Theory," *Journal of Economic Literature*, 12, 1255–1289.

Shapiro, B. (2019): "Promoting Wellness or Waste? Evidence from Antidepressant Advertising," *Working paper*.

Shapiro, B. T. (2018): "Positive Spillovers and Free Riding in Advertising of Prescription Pharmaceuticals: The Case of Antidepressants," *Journal of Political Economy*, 126.

Silverman, J. J., M. Galanter, M. Jackson-Triche, D. G. Jacobs, J. W. Lomax, M. B. Riba, L. D. Tong, K. E. Watkins, L. J. Fochtmann, R. S. Rhoads, and J. Yager (2015): "The American Psychiatric Association Practice Guidelines for the Psychiatric Evaluation of Adults," *American Journal of Psychiatry*, 172, 798–802.

Sinkinson, M. and A. Starc (2019): "Ask Your Doctor? Direct-to-Consumer Advertising of Pharmaceuticals," *The Review of Economic Studies*, 86, 836–881.

Song, M., S. Nicholson, and C. Lucarelli (2017): "Mergers with Interfirm Bundling: A Case of Pharmaceutical Cocktails," *The RAND Journal of Economics*, 48, 810–834.

Ventola, C. L. (2011): "Direct-to-Consumer Pharmaceutical Advertising," *Pharmacy & Therapeutics*, 36, 10.

Wakamori, N. (2015): "Portfolio Considerations in Differentiated Product Purchases: An Application to the Japanese Automobile Market," *Discussion Paper Series of SFB/TR 15 Governance and the Efficiency of Economic Systems 499*.

Williams, S. Z., G. S. Chung, and P. A. Muennig (2017): "Undiagnosed Depression: A Community Diagnosis," *SSM - Population Health*, 3, 633–638.

A Appendix

A.1 Proof of the Proposition on Hidden Markov Models

Proposition. *Let f and g stand for any of the values that the first-order Markov chain ψ_{it} can take. Define the joint probability of the observed sequence up to time t and the Markov chain being in state f at time t, conditional on all observables $\mathbf{X_i}$, the unobservable $\nu_\mathbf{i}$, and the parameters of the model Θ:*

$$\phi_{it}(f) = Pr(c_{i1}, \ldots, c_{it}, \psi_{it} = f | \Theta, \mathbf{X}_i, \nu_i)$$

Claim: $\phi_{it}(f)$ can be computed recursively as:

$$\begin{aligned}
\phi_{i1}(f) &= \pi_f s^f_{ic_{i1}1} && \text{for } f \in \{H,U\} \\
\phi_{it}(f) &= \Big[\sum_{g\in\{H,U\}} \phi_{it-1}(g)\pi_{gf}\Big] s^f_{ic_{it}t} && \text{for } f \in \{H,U\} \text{ and } t \in \{2,\ldots,\mathcal{T}\}
\end{aligned}$$

Proof. By definition

$$\begin{aligned}
\phi_{i1}(f) &= \pi_f s^f_{ic_{i1}1} \\
&= Pr(\psi_{i1} = f) Pr(c_{i1}|\psi_{i1} = f, \Theta, \mathbf{X}_i, \nu_i) \\
&= Pr(c_{i1}, \psi_{i1} = f|\Theta, \mathbf{X}_i, \nu_i)
\end{aligned}$$

Thus, $\phi_{it}(f)$ is the joint probability of choosing the observed choice c_{i1} and being in health state f at $t = 1$.

For any $t > 1$:

$$\begin{aligned}
\phi_{it}(f) &= \Big[\sum_{g\in\{H,U\}} \phi_{it-1}(g)\pi_{gf}\Big] s^f_{ic_{it}t} \\
&= \Big[\sum_{g\in\{H,U\}} Pr(c_{i1},\ldots,c_{it-1},\psi_{it-1} = g|\Theta,\mathbf{X}_i,\nu_i) Pr(\psi_{it} = f|\psi_{it-1} = g)\Big] s^f_{ic_{it}t} \\
&= \Big[\sum_{g\in\{H,U\}} Pr(c_{i1},\ldots,c_{it-1},\psi_{it-1} = g|\Theta,\mathbf{X}_i,\nu_i) Pr(\psi_{it} = f|\psi_{it-1} = g, c_{i1},\ldots,c_{it-1},\Theta,\mathbf{X}_i,\nu_i)\Big] s^f_{ic_{it}t} \\
&= \Big[\sum_{g\in\{H,U\}} Pr(c_{i1},\ldots,c_{it-1},\psi_{it-1} = g,\psi_{it} = f|\Theta,\mathbf{X}_i,\nu_i)\Big] s^f_{ic_{it}t} \\
&= Pr(c_{i1},\ldots,c_{it-1},\psi_{it} = f|\Theta,\mathbf{X}_i,\nu_i) Pr(c_{it}|\psi_{it} = f,\Theta,\mathbf{X}_i,\nu_i) \\
&= Pr(c_{i1},\ldots,c_{it-1},\psi_{it} = f|\Theta,\mathbf{X}_i,\nu_i) Pr(c_{it}|\psi_{it} = f,c_{i1},\ldots,c_{it-1},\Theta,\mathbf{X}_i,\nu_i) \\
&= Pr(c_{i1},\ldots,c_{it-1},c_{it},\psi_{it} = f|\Theta,\mathbf{X}_i,\nu_i)
\end{aligned}$$

Moving from line 2 to 3 is possible because of the first-order Markov assumption. Moving from line 4 to 5 requires integrating out ψ_{it-1}. Moving from line 5 to 6 stems from the assumption that there is no structural dependence in choices—this assumption, however, is not required for the proof of the proposition and can be relaxed. The crucial assumption is the first-order Markov assumption. □

A.2 Alternative Assumptions on the Unobservables

Table 12: First Stage (MSL) Estimates - Sensitivities

	(1)/(2) Main/No Spillovers		(3) Uncorrelated Preferences		(4) No Health Shocks	
	Est	SE	Est	SE	Est	SE
OOP	-0.206	(0.143)	0.077	(0.141)	-0.284***	(0.102)
×MDD	0.721***	(0.083)	1.139***	(0.083)	0.709***	(0.068)
×Depression-Other	0.039	(0.093)	-0.034	(0.093)	0.358***	(0.082)
×Female	-0.041	(0.101)	-0.353***	(0.106)	-0.251***	(0.077)
×Age	0.103**	(0.048)	0.164***	(0.052)	0.079**	(0.038)
×Salaried	0.221	(0.14)	0.002	(0.135)	0.001	(0.091)
×Union	0.283**	(0.136)	0.684***	(0.113)	-0.22*	(0.124)
Demographics×Generic						
Female	0.958***	(0.095)	1.169***	(0.139)	1.504***	(0.06)
Age	0.363***	(0.046)	0.502***	(0.063)	0.345***	(0.031)
Age^2	-0.174***	(0.051)	-0.047	(0.065)	0.222***	(0.032)
HMO	0.277	(0.181)	0.077	(0.19)	-0.627***	(0.121)
POS	-0.229	(0.198)	-0.5**	(0.247)	-0.86***	(0.13)
PPO	0.518***	(0.161)	0.019	(0.165)	-0.057	(0.105)
Salaried	-0.574***	(0.133)	-0.626***	(0.146)	-0.876***	(0.085)
Hourly	-1.099***	(0.126)	-0.482***	(0.152)	-1.424***	(0.089)
Union	1.148***	(0.144)	0.929***	(0.148)	1.132***	(0.104)
Non-Union	1.079***	(0.125)	0.282**	(0.119)	0.599***	(0.082)
Full-Time	-0.347***	(0.117)	-0.512**	(0.23)	-0.467***	(0.076)
MDD	0.487***	(0.066)	1.022***	(0.065)	2.678***	(0.047)
Depression-Other	0.662***	(0.078)	1.095***	(0.073)	3.489***	(0.066)
Demographics×Branded						
Female	1.106***	(0.142)	0.76***	(0.136)	2.942***	(0.108)
Age	0.132*	(0.08)	-0.355***	(0.076)	0.228***	(0.051)
Age^2	-0.066	(0.058)	-0.993***	(0.058)	-0.057	(0.037)
HMO	-0.254	(0.205)	0.062	(0.213)	-1.142***	(0.149)
POS	-0.151	(0.219)	0.461**	(0.229)	-0.639***	(0.158)
PPO	1.232***	(0.189)	0.58***	(0.191)	0.931***	(0.131)
Salaried	-0.114	(0.206)	-1.173***	(0.199)	-0.559***	(0.148)
Hourly	-0.571***	(0.141)	0.371***	(0.138)	-1.046***	(0.101)
Union	0.971***	(0.204)	1.139***	(0.205)	2.419***	(0.161)
Non-Union	1.29***	(0.13)	0.336***	(0.123)	0.821***	(0.092)
Full-Time	0.066	(0.137)	-0.262**	(0.133)	-0.827***	(0.096)
MDD	0.301**	(0.117)	1.32***	(0.118)	2.685***	(0.096)
Depression-Other	1.722***	(0.149)	3.04***	(0.156)	4.281***	(0.118)
Demographics×Therapy						
Female	1.134***	(0.134)	0.722***	(0.138)	1.616***	(0.098)
Age	-0.202***	(0.068)	-0.298***	(0.074)	-0.483***	(0.049)
Age^2	-0.144***	(0.041)	-0.36***	(0.044)	-0.266***	(0.029)
HMO	1.025***	(0.233)	0.331*	(0.193)	0.058	(0.146)
POS	1.456***	(0.223)	1.09***	(0.203)	0.791***	(0.147)
PPO	1.779***	(0.212)	0.893***	(0.182)	0.855***	(0.133)
Salaried	-1.008***	(0.206)	-0.596***	(0.197)	-0.098	(0.129)
Hourly	-0.706***	(0.116)	0.268**	(0.107)	-0.296***	(0.078)
Union	0.784***	(0.178)	0.159	(0.169)	0.044	(0.156)
Non-Union	0.841***	(0.103)	-0.366***	(0.102)	-0.241***	(0.073)
Full-Time	-0.668***	(0.101)	-0.465***	(0.109)	-1.206***	(0.085)
MDD	0.335***	(0.102)	0.267***	(0.1)	2.042***	(0.086)
Depression-Other	-0.589***	(0.135)	-0.353***	(0.128)	0.695***	(0.107)
Log-likelihood	-31,839.5		-31,927.2		-37,952.4	
Observations	445,500		445,500		445,500	

Notes: Bundle-MSA-month fixed effects included. Demographics are interacted with product dummies, OOP and OOP interactions are not. For the multiproduct bundles (not shown), the demographic effects equal the sum of the demographic effects on the component single-product bundles. OOP prices and age are measured in standard deviations ($24.52 and 10.62 years, respectively) and centered around their means ($26.70 and 44.85 years). Employees can be salaried, hourly, or unknown; unionized, non-unionized, or unknown; full-time, early retiree, part-time, or "other." Depression diagnosis can be major depressive disorder (MDD), dysthymia, or depression-other. The omitted health insurance type is CDHP/HDHP. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

Table 13: Covariance/Correlation Matrix of Unobserved Preferences - Sensitivities

(1) Main / (2) No Spillovers

	Generic	Branded	Therapy
Generic	8.64	0.34	0.25
Branded	4.06	16.25	0.44
Therapy	2.34	5.51	9.84

(3) Uncorrelated Preferences

	Generic	Branded	Therapy
Generic	7.37	0	0
Branded	0	19.85	0
Therapy	0	0	7.98

(4) No Health Shocks

	Generic	Branded	Therapy
Generic	22.03	0.89	0.73
Branded	23.84	32.21	0.76
Therapy	13.59	17.19	15.84

Notes: The first stage of the estimation is the same for the models with and without advertising spillover effects. Variance and covariance terms on the main diagonal and below; correlation coefficients—above. All parameter estimates are significant at the 1% significance level.

Table 14: Markov Chain Health Shock Transition Probabilities and Stationary Distribution - Sensitivities

(1) Main / (2) No Spillovers

Panel A: Transition Probabilities

	Healthy_{t+1}	Depressed_{t+1}
Healthy_t	0.9982	0.0018
	(0.0000)	(0.0000)
Depressed_t	0.0175	0.9825
	(0.0000)	(0.0000)

Panel B: Stationary Distribution

	Healthy	Depressed
Long-run share	0.9066	0.0934

(3) Uncorrelated Preferences

Panel A: Transition Probabilities

	Healthy_{t+1}	Depressed_{t+1}
Healthy_t	0.9984	0.0016
	(0.0000)	(0.0000)
Depressed_t	0.0159	0.9841
	(0.0000)	(0.0000)

Panel B: Stationary Distribution

	Healthy	Depressed
Long-run share	0.9066	0.0934

Note: Standard errors in parentheses. The first stage of the estimation is the same for the models with and without advertising spillover effects. The model with no health shocks has no transition probabilities and does not appear in the table.

Table 15: Second Stage, OLS Estimates - Sensitivities

	(1) Main	(2) No Spillovers	(3) Uncorr. Pref.	(4) No Health Shocks
Generic: Ads	-0.0026***		-0.0027***	-0.0012***
	(0.0006)		(0.0006)	(0.0004)
Branded: Ads	0.0059***	0.0062***	0.0073***	0.005***
	(0.0007)	(0.0007)	(0.0007)	(0.0005)
Therapy: Ads	0.0009		0.0011	0.001**
	(0.0006)		(0.0007)	(0.0005)
Time Effects	Year-Trend	Year-Trend	Year-Trend	Year-Trend
Market FEs	MSA	MSA	MSA	MSA

Notes: Standard errors in parentheses. For the multiproduct bundles, the advertising effects equal the sum of the effects on the component single-product bundles. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 16: Second Stage, 2SLS Estimates - Sensitivities

	(1) Main	(2) No Spillovers	(3) Uncorr. Pref.	(4) No Health Shocks
Panel A: average complementarity				
Generic-Therapy	-0.1939***	-0.1939***	0.5424***	0.6883***
	(0.0113)	(0.0113)	(0.0113)	(0.0084)
Branded-Therapy	-0.7716***	-0.7716***	0.7139***	0.2379***
	(0.0124)	(0.0124)	(0.0128)	(0.0094)
Panel B: advertising effects on each bundle				
Generic: Ads	0.0053***		0.0053***	0.0039***
	(0.0025)		(0.0024)	(0.0016)
Branded: Ads	0.0072***	0.0096***	0.0076***	0.0048***
	(0.003)	(0.0028)	(0.0031)	(0.0021)
Therapy: Ads	0.0071***		0.0068***	0.0036***
	(0.0027)		(0.0028)	(0.002)
Time Effects	Year-Trend	Year-Trend	Year-Trend	Year-Trend
Market FEs	MSA	MSA	MSA	MSA
Panel C: first stage of instrumental variables regression				
Adj Avg Price per Ad			-0.0057	
Partial F-stat			504.8	

Notes: Standard errors in parentheses. For the multiproduct bundles, the advertising effects equal the sum of the effects on the component single-product bundles. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 17: Demand Elasticities with Respect to Advertising, Sensitivity

	(1) Main	(2) No Spillovers	(3) Uncorr. Pref.	(4) No Health Shocks
		Bundles		
Outside Option	-0.0053	-0.0016	-0.0053	-0.0034
Generic-Only	0.0033	-0.0016	0.0033	0.0642
Branded-Only	0.0063	0.0138	0.0070	0.0797
Therapy-Only	0.0061	-0.0016	0.0057	0.0583
Generic-Therapy	0.0147	-0.0016	0.0142	0.1259
Branded-Therapy	0.0177	0.0138	0.018	0.1414
		Products		
Outside Option	-0.0053	-0.0016	-0.0053	-0.0034
Generic	0.0050	-0.0016	0.0051	0.0654
Branded	0.0076	0.0138	0.0089	0.0805
Therapy	0.0093	-0.0004	0.0097	0.0613

Note: Demand elasticities with respect to branded drug advertising are calculated by averaging individual-level elasticities over patients and time periods.