RAND Journal of Economics
Vol. 42, No. 2, Summer 2011
pp. 205–245

Advertising, the matchmaker

Bharat N. Anand*

and

Ron Shachar**

We empirically study the informational role of advertising in matching consumers with products when consumers are uncertain about both observable and unobserved program attributes. Our focus is on the network television industry, in which the products are television shows. We estimate a model that allows us to distinguish between the direct effect of advertising on utility and its effect through the information set. A notable behavioral implication is that exposure to informational advertising can decrease the consumer's tendency to purchase the promoted product. The structural estimates imply that an exposure to a single advertisement decreases the consumer's probability of not choosing her best alternative by approximately 10%. Our results are relevant for industries characterized by product proliferation and horizontal differentiation.

1. Introduction

■ Product proliferation is occurring in virtually every sector today.[1] Among other things, an important consequence of proliferation is to increase consumer uncertainty about *observable* product attributes. For example, what is Salman Rushdie's new treatise about? Does the new organic store on the street corner carry regular Coke? Can Toyota's Prius run on electricity when it runs out of gas? Does Apple's iPhone have the capability to record video or replay music through the radio? In contrast to experience attributes like the taste of Coca-Cola or the quality of a haircut, information about these attributes can be entirely learned through costly search. Alternatively, as recognized by Nelson (1974) and formalized by Grossman and Shapiro (1984), consumers can obtain relevant information about observable attributes from advertising content,

*Harvard University; banand@hbs.edu.
**Arison School of Business at IDC, Israel; ronshachar@idc.ac.il.
We are deeply grateful to Dmitri Byzalov for excellent research assistance, and to participants in seminars at Carnegie-Mellon University, Columbia University, Econometrics in Tel-Aviv, Haifa University, Harvard University, Hebrew University, MIT, NBER IO Meetings, NBER Productivity Seminar, Northwestern University, NYU, Oxford University, Stanford University, Tel-Aviv University, University of Pennsylvania, University of Rochester, University of Toronto, University of Washington, and Yale University, for valuable comments. Anand is grateful to the Division of Research at Harvard Business School for financial support. Part of Shachar's work was done while he was a faculty member in Tel Aviv University.

[1] In the United States, the average supermarket carries 275 varieties of cereal, the average grocery store carries over 600 varieties of cold medicine, and the number of mutual funds offered is over 13,000. Dramatic increases in variety are also seen for products ranging from cars to computers and telephone calling plans.

thereby improving matches between consumers and products. Indeed, the explosion of more targeted forms of advertising in recent years suggests that this "matchmaking" role of advertising should only increase.

In this article, we empirically study the informational role of advertising in matching consumers with products. We examine this phenomenon within the context of the network television industry. Advertising is the main revenue stream for networks, and television advertising accounts for over 40% of all advertising in the economy. Roughly one sixth of all television advertising is in the form of "tune-ins"—networks promoting their own shows—resulting in high advertising/sales ratios for TV firms. Our focus here is on television tune-ins (also referred to as "cross-promotions" or "previews"), and how they affect the matching of viewers and programs. The findings are relevant not only to TV but to other industries characterized by product proliferation and horizontal product differentiation. Furthermore, understanding how advertising works can also shed light on broader questions related to the future of advertising revenue streams in the TV industry itself.[2]

To date, there remains little prior empirical work on the matching role of advertising. Whereas theory has examined the role of advertising in markets with consumer uncertainty about observable product attributes (e.g., Grossman and Shapiro, 1984), empirical work has largely focused on markets with experience products. Certain papers in this latter stream also make important advances in tackling a familiar challenge that confronts any empirical analysis of advertising—that the informational effect and direct effect appear to be observationally equivalent. They do so by using individual-level data on choices to exploit various sources of identification for experience products. For example, Ackerberg (2001, 2003) exploits variation in advertising effectiveness across consumers with different levels of experience about a product in order to separate the impact of advertising on "experience information" from its direct effect. Similarly, Erdem and Keane (1996) identify ads' precision via the impact of advertising exposures on variation in consumer choices over time (i.e., consumers who were exposed to more advertisements have better information and are therefore more likely to persist with the most suitable alternative). Central to the empirical implementation in these papers is that an individual is observed on multiple purchase occasions for the same product.

Like these studies, our data contain repeated observations on consumers (viewers). Unlike them, each "time slot" corresponds to a different product (TV program), and therefore we do not observe repeated choices for the same product by an individual. Even without such data, we can separate the informative and direct effects of advertising. An important reason is that the informative role of ads is somewhat different in our model—we focus on the role of advertising in providing information on observable product attributes, rather than (as in the case of prior work) on unobserved consumer preferences.[3] This provides sources of identification beyond those in prior work. For example, our data contain information on observed attributes of programs (x), viewer demographics (z), and the number of advertisements that each viewer was exposed to for each program. The match between consumers and products is a function of xz (the interaction of program attributes and viewer demographics). Whereas this match is observed by the researcher, it is not observed by viewers who are uncertain about x. Correlation between the match and choices yields a measure of how informed viewers are about programs. Viewers who are uninformed about program characteristics will randomly pick programs and thus the correlation between their match and their choices is low; for fully informed viewers, the correlation will be high. Now, variation in this correlation across viewers who were exposed to different numbers of advertisements for the

[2] For example, advertising is threatened by the rise of devices such as Replay TV and TiVo that both make it easier to skip advertisements and easier to target advertisements to viewers. Consumers' decision to skip advertisements or not depends on whether they consider ads a nuisance that "persuades" (or "brainwashes") them into purchasing products they do not want, or view them as providing useful information about products they want, thereby resulting in improved matching.

[3] By "observable attributes," we refer throughout to product attributes that are observed to the researcher. Consumers may be uncertain about these attributes.

relevant program identifies the informative role of advertising. The larger the correlation for viewers who were exposed to more advertisements, the larger the "matchmaking role" of advertising.

Another way to see this is to consider the following simple example. Suppose the choice data reveal that men like to watch action dramas but dislike romantic dramas, whereas women have opposite preferences. Then, if the matching role were present, it follows that men who are exposed to more advertisements for action dramas are more likely to view such programs, whereas exposure to advertisements for romantic dramas makes them *less* likely to view those programs. Furthermore, and importantly, the reverse patterns hold for women.

This argument reveals how identification is feasible even without data on individuals' repeated choices over a product. Beyond this, the logic of identification outlined here has two interesting implications. First, identification does not rely on assuming that the direct effect of advertising is similar across viewers (for example, the direct effect could be different for women and men in the example above, and even different across programs, as we discuss later). Second, the matching role implies a "consumption-deterrence" property of advertising: namely, advertising exposure can decrease a viewer's propensity to watch a program.

The example above is obviously stylized. There are various additional features—unobserved attributes, unobserved heterogeneity in preferences, state dependence, and endogenous advertising exposures—that we explicitly integrate into the model and account for in the structural estimation. These issues are described in detail later. However, the source of identification described here, namely that advertising improves the realized observed match between consumers and products, remains in the presence of all these additional features. Furthermore, there are additional sources of identification in the data (for example, the role of the multiproduct firms' profile) that assist in the empirical implementation, and that we describe later.

Our data contain information on consumption and exposures to advertising at the individual level. To create these data, we obtained Nielsen individual-level panel data on television-viewing choices for one week in November 1995. We created data on show attributes, and recorded all the advertisements for these television shows (also called "previews") that were aired during that week. Combining our records with the Nielsen panel data gives us the required data to estimate the model. Although, as mentioned earlier, our data do not contain information on repeated product choices for an individual, there are certain advantages over typical data sets and settings used for estimating advertising models. First, accounting for the cost of leisure in consumption, television shows are clearly one of the most important consumption products.[4] Second, the monetary cost of viewing a network television show is zero, and the nonmonetary cost is the same (for each individual) across shows in any period. This avoids the problem of unobserved variation of prices due to the availability of coupons. Third, almost all the commercials for television shows appear on TV. This enables us to create a comprehensive data set of exposures to advertisements.

Last, an important feature of our data set is that it enables us to overcome the well-known endogeneity problem of advertising exposures. Advertising spending and targeting are chosen strategically by firms. Thus, the ad exposure variable depends on the unobservables of the model. Not accounting for the potential correlation between ad exposures and the unobservables would result in inconsistent estimates. To resolve the endogeneity problem, our likelihood function includes the joint distribution of ad exposures and the unobservables. In most applications, constructing this joint distribution requires additional ad hoc assumptions, and requires collecting additional data on variables that determine ad exposures. In our application (television-viewing choices), the joint distribution comes directly from the model and the data. The reason is that ad exposures are a function of previous consumption (viewing) choices, which are already part of the model. A detailed discussion of the endogeneity problem and its solution is presented in Section 4.

[4] The Television Bureau of Advertising (www.tvb.org) reports that in 2001 the average adult man in the United States spent 4 hours and 19 minutes watching television per day, and the average woman spent 4 hours and 51 minutes per day.

Although the aim of this study is to structurally estimate the parameters of the model, we start our empirical investigation with a nonstructural test of the model's implications. Indeed, we find that the match experienced by the individual is a positive function of the number of advertisements to which she was exposed.[5]

The parameter of interest in the structural estimation is the precision (reciprocal of the variance) of the noisy advertising signal. If the estimate of this parameter were equal to zero, then advertising does not have any informational role. In our data, the estimate of the precision of advertising signals is positive and statistically different from zero at the 1% significance level. Furthermore, the behavioral impact of advertising signals is substantial. For example,we find that the precision of one advertisement is equal to the precision of all the other product-specific signals together.

The structural estimates are used to illustrate that the matching role of advertising is significant, in several ways. For example, it is shown that an exposure to one advertisement decreases a viewer's probability of watching a show, which yields a relatively low match from 0.355 to 0.184. Furthermore, we find that an exposure to a single advertisement decreases the consumer's probability of not choosing her best alternative by approximately 10%.

For each product, a firm obviously intends its advertisements to reach consumers whose response to exposures is the largest. The structural estimates serve to locate the advertising placements that maximize the networks' profits, and can therefore be used to examine the targeting strategies of the television networks. It turns out that some simple general rules (e.g., placing an advertisement for a show in the preceding one) characterize these "optimal" locations. Furthermore, the data suggest that the networks employ similar rules. This exercise and its results can be viewed as a nonformal specification test of the model.

The key ingredients of the model are not industry specific. However, applying the model to the television example requires accounting for the specificity of viewing behavior. Thus, in order to make the presentation clear, we start by describing the data, in Section 2. Section 3 presents the model and its implications; Section 4 discusses estimation issues; Section 5 presents the results; and Section 6 examines the normative and positive implications of the theory. Section 7 concludes. The next subsection reviews the related literature.

□ **Related literature.** The availability of individual-level data on consumption and advertising exposures has spawned an empirical literature on this topic (see Bagwell, 2003). Most studies include ad exposures as an element in the utility function.[6] Some (notably the influential papers by Ackerberg, 2003 and Erdem and Keane, 1996) allow advertising to affect the information set. Including advertising in the information set has two justifications. First, advertising can directly convey information on products' existence and attributes (Butters, 1977; Grossman and Shapiro, 1984; Anderson and Renault, 2005). Second, advertising can signal product quality in equilibrium (Nelson, 1974; Kihlstrom and Riordan, 1984; Milgrom and Roberts, 1986). Whereas in the first approach advertising content is informative, the second approach assumes that ad content is empty and that consumers make inferences about product attributes from firms' actions (regarding advertising intensity, for example). Furthermore, the focus of the "directly informative effect" is on resolving uncertainty about horizontal product attributes, whereas the "indirect signaling effect" of advertising is concerned with a product's vertical attribute (i.e., quality). As a result of these differences, these two approaches also have different behavioral implications.

Thus far, most empirical studies that have included advertising in the information set have focused on the signalling theory.[7] Our focus in this article is on the first mechanism of informative advertising—that is, information directly conveyed through advertising content.

[5] In this test, as in other nonstructural examinations that are reported in Section 5, we control for the direct effect of advertising on consumers' utility.

[6] For example, see Roberts and Samuelson (1988), Stern and Trajtenberg (2001), Nevo (2001), and Coscelli and Shum (2004).

[7] See also Horstmann and Macdonald (2001).

Empirical analyses of how this matching role of advertising impacts consumer choices has remained somewhat unexplored until now.[8] This is surprising, because advertising is often for horizontally differentiated products and the content of many ads is informative. This study aims to fill this gap.

Grossman and Shapiro (1984; hereafter GS) were the first to theoretically examine the role of advertising in matching consumers with products. Because of the need to take theory to the data, the model presented here differs from GS in several ways. For example, advertising need not convey full and accurate information about product attributes, and consumers may have other sources of information besides advertising. Consumer preferences are formulated here in the spirit of Lancaster (1971) rather than the circular city model of Salop (1979) followed by GS. These differences in approach not only facilitate empirical estimation but also generate differences in theoretical predictions (including the consumption-deterring aspect of advertising).[9]

The closest empirical papers to ours are Ackerberg (2003) and Erdem and Keane (1996).[10] Like those studies, we examine the effect of ads through the information set, exploit information on individual-level choices, and estimate a structural model with Bayesian learning. Furthermore, part of our identification also rests on the structure that the model imposes on the variance of choices by individuals (this is the central source of identification in their work). The central difference, as described earlier, is that those studies focus on uncertainty about (unobserved) consumer preferences, whereas we examine uncertainty about observable product attributes. Put differently, the typical (if unstated) presumption in prior studies is that consumers have (at least some) uncertainty about preferences for idiosyncratic features of the product and learn those tastes through use and/or advertising. The model here is different, however. Consumers are assumed to have fixed preferences for product characteristics, but are repeatedly faced with products with new characteristics, about which they can learn through advertising (as well as other miscellaneous sources). Economically, the relevance of each model will vary across product markets, as noted earlier. Methodologically, this difference in focus has at least two important implications. First, even without data on repeated choices, we can identify the informative effect of advertising. Second, our interest in the matching role highlights the consumption-deterring role of advertising. This property is embedded in certain previous models as well but, as far as we know, has neither been explicitly identified nor exploited in prior work.[11,12]

2. Data and preliminary evidence

■ The first part of this section presents the data and their suitability for testing our advertising theory. The second part offers preliminary evidence on the matching role of advertising.

□ **Data.** The data consist of information about the attributes of products and the characteristics, choices, and advertising exposures of consumers in the U.S. television industry. The television example fits nicely into the setting of the model (described briefly above and in detail in the following section): products are differentiated, consumers are heterogeneous, consumers are

[8] Milyo and Waldfogel (1999) have examined some effects of informative advertising on market equilibrium without observing consumer choices. Mitra and Lynch (1995) examine the matching role of advertising in an experimental context.

[9] In GS, any exposure to an advertisement increases an individual's tendency to purchase the promoted product. This is because, in their setting, a consumer who is not exposed to advertisements is ignorant about the existence of this firm, and thus her probability of purchasing such a product is zero.

[10] More broadly, our focus on direct information disclosure, rather than on signalling through actions, is related as well to Jin and Leslie (2003) and Chernew, Gowrisankaran, and Scanlon (2008).

[11] Beyond this, there are other differences as well. First, in contrast to Erdem and Keane (and like Ackerberg), we include advertising both in the information set and directly in the utility. Second, we do not treat advertising exposures as exogenous.

[12] A notable exception is Johnson and Myatt's (2006) analysis of promotional hype versus real information in advertising. In contrast to promotional hype, "real information allows a consumer to learn of his personal match with the product's characteristics" and rotates, rather than shifts, the demand curve as a result. Their taxonomy therefore accomodates the possibility that "supplying real information may sometimes lower rather than raise sales (and profits)."

uncertain about product attributes, and some of these attributes are observable.[13] Before presenting the data, we discuss consumers' uncertainty in our application.

There are several reasons why consumers may not be fully informed about attributes of television shows. First, there are frequent changes in the weekly schedule. The most dramatic change takes place at the beginning of the season, when most of the shows are either new ones or veterans aired in a new time slot. For this reason, we have requested data from an early stage of the season (about five weeks into it). Second, episodes constantly evolve for each program and the focus of a show frequently shifts from one episode to another. For example, one episode might focus on a female character and her personal dilemmas, whereas the next is centered on her male spouse. Last, viewers are unlikely to remember the schedule precisely. Thus, even if they knew the attributes of all the shows, they would be uncertain about what a firm might offer in a particular time slot and hence uncertain about show attributes. Indeed, using the same data set, we have shown earlier (Anand and Shachar, 2004) that viewers are uncertain about product attributes.[14]

The data on individual characteristics and choices were obtained from A.C. Nielsen, and the rest of the data were designed and created for the purpose of this study. The viewing data cover prime-time programming (8:00–11:00 p.m.) of the four national broadcast networks (ABC, CBS, NBC, and Fox) for November 6–11, 1995.

The data sets are presented in the following order: product attributes, consumer characteristics, consumption choices, and exposures to advertisements.

Product (show) characteristics. The four networks aired 64 shows during these five days. None of the shows was a rerun. We coded the show attributes based on prior knowledge, publications about the shows, and viewing each one of them. Following previous studies of viewing choices that identified the importance of observable show attributes, we categorize shows based on their genre and cast demographics. Rust and Alpert (1984) present five show categories—comedies and action dramas, for example—and show that viewers differ in their preferences over these categories. We use the following categories: *situational comedies*, also called "sitcoms" (31 shows fall into this category), *action dramas* (16 shows), *romantic dramas* (9 shows), *news magazines*(6 shows), and *sports events* (2 shows).

Shows were also characterized by their cast demographics. Shachar and Emerson (2000) suggest that *similarity* between the individual and the cast of a show plays an important role in viewing choices because viewers enjoy a show more when they can *identify* with the main characters. Using a measure of similarity that is based on the demographic match between an individual and a show's cast, they demonstrate that the data support their hypothesis. For example, younger viewers tend to watch shows with a young cast, whereas older viewers prefer an older cast. We use the following categories: *Generation-X*, if the main characters in a show are older than 18 and younger than 34 (21 shows fall into this category); *Baby Boomer*, if the main show characters are older than 35 and younger than 50 (12 shows); *Family*, if the show is centered around a family (11 shows); *African-American* (7 shows); *Female* (15 shows); and *Male* (22 shows).[15]

Consumer characteristics and choices (the Nielsen data). We obtained data on individuals' viewing choices and characteristics from Nielsen Media Research, which maintains a sample of over 5000 households nationwide.[16] Nielsen installs a People Meter (NPM) for *each* television

[13] The empirical strategy relies on the attributes of the product that are observable and verifiable without experience.

[14] There are, obviously, differences among viewers in how much they know—some of them might read *TV Guide*, others not. The model allows viewer heterogeneity in information.

[15] It turns out that even though a show may have several main characters, in no case do these characters belong to different age categories. As a result, every show is allocated to a single age category. The same is true for the other cast demographics. Note that in the case of *Friends*, because the main characters are both female and male, the show was not allocated to any gender category.

[16] Using 1990 Census data, the sample is designed to reflect the demographic composition of viewers nationwide. The sample is revised regularly, ensuring, in particular, that no single household remains in the sample for more than two years.

TABLE 1 Individual Observable Characteristics: Definitions and Summary Statistics

Variable	Definition	Mean	Standard Deviation
Teens	Viewer is between 6 and 17 years old (in November 1995)	0.1421	0.3491
Gen − X	Viewer is between 18 and 34 years old (in November 1995)	0.2400	0.4272
Boom	Viewer is between 35 and 49 years old (in November 1995)	0.2764	0.4474
Older	Viewer is older than 50 years	0.3415	0.4742
Female	Female viewer	0.5319	0.4991
Family	Viewer lives in a household with (according to Nielsen codes) a "woman of the house" (i.e., female over the age of 18) present	0.4304	0.4953
Income	Measured on unit interval, where the limits are: zero if the income is less than $10,000, and one if the income is $40,000 and over	0.8333	0.2259
Education	Measured on unit interval, where the limits are: zero if the years of school are less than 8, and one if it is 4 or more years college	0.7421	0.2216

set in the household. The NPM records the channel being watched on each television set, and a special remote control records the individuals watching each TV. Thus, the viewing choices are individual specific. Although criticized occasionally by the networks, Nielsen data still provide the standard measure of ratings for both network executives and advertising agencies.

Although the NPM is calibrated for measurements each minute, the data available to us provide quarter-hour viewing decisions, measured as the channel being watched at the midpoint of each quarter-hour block. Thus, we observe viewers' choices in 60 time slots. This study confines itself to East Coast viewers, to avoid problems arising from ABC's Monday night programming.[17] Finally, viewers who never watched television during weeknight prime time and those younger than six years of age are eliminated from the sample. From this group, we randomly selected individuals with a probability of 50% (in order to decrease the computational burden). This gives us a final sample of 1675 individuals. On average, at any point in time, only 25% of the individuals in the sample watch network television.

In addition to viewer choices, Nielsen also reports their personal characteristics. Our data include the age and the gender of each individual, and the income, education, cable subscription, and county size for each household. Table 1 defines the variables created based on this information, as well as their summary statistics.

Data on exposures to advertising. We taped all the shows for the four networks during the week that started on November 6, 1995. We then coded the appearance of each advertisement for the television shows. For example, on Monday at 9:10 p.m., there was an advertisement for the ABC news magazine *20/20* (this show aired on Friday at 10:00 p.m.). This information was matched with the Nielsen viewing data to determine an individual's exposure to advertisements. For example, an individual who watched ABC on Monday at 9:10 p.m. was exposed to the advertisement mentioned above. Summing over all time slots, we get the number of exposures of individual i with respect to each show in the week. In 1995, these advertisements, also referred to as "promos," usually included the broadcast time of the show and clips from the actual episode.

Because our Nielsen viewing data start on Monday, we cannot determine the exposure to advertisements that were aired before that day. This means that our data miss some ad exposures. This problem obviously affects the exposure variable for shows that were broadcast on Monday and Tuesday but does not seem to influence those that aired on Wednesday through Friday. Thus, in the nonstructural tests, we use only the data for Wednesday through Friday, and in the structural estimation, we allow the advertising parameters to differ across these two parts of the week.

[17] ABC features Monday Night Football, broadcast live across the country; depending on local starting and ending times of the football game, ABC affiliates across the country fill their Monday night schedule with a variety of other shows. Adjusting for these programming differences by region would unnecessarily complicate this study.

For the Wednesday through Friday shows, the mean number of advertisements aired per show is 4.9, and the median is 4. On average, an individual who watched television on Monday or Tuesday was exposed to 0.56 advertisements for each show on Wednesday through Friday.

□ **Preliminary evidence on matching.** In order to test the matching role of advertising, one needs a precise measure of the match between consumer tastes and product attributes. The model, presented in the next section, introduces such a measure. However, before proceeding to the model, it is useful to examine whether the data appear to support a matching role for advertising.

A simple measure of the match between consumers and products is based on the match between an individual's demographics and those of a show's cast. The match variable, *Match*, is based on three demographic characteristics: age, gender, and family status. It counts the number of characteristics that are identical for both the show and the individual. For example, for a Generation-X single female viewer and a Generation-X show with a single, male cast, $Match = 2$.[18]

The matching hypothesis is that exposure to advertising increases the consumer's familiarity with shows and, thus, the likelihood that she *chooses* a show that is suitable for her. We evaluate a consumer's choice by the resulting gain in the match variable, *Match*. For example, for a consumer who chose show 1 over show 2, the gain is $Match_1 - Match_2$. The gain can, obviously, be negative when the consumer chooses a show whose attributes yield a lower match for her than the competing show or shows. The matching hypothesis is that the gain in the match variable increases in the number of ad exposures. This hypothesis is based on the effect of advertising on choices through the information set.

Another challenge that we face in executing this test is that an increase in the number of ad exposures might affect choices not only through the information set but also via a direct effect on the utility. In the structural estimation, we estimate both effects simultaneously. Here we focus on the informative role of advertising by "cancelling out" the direct effect. We do so by focusing only on individuals who are exposed to the same number of ads for each of the competing shows in a time slot.[19] For such viewers, the direct effect is the same across all alternatives. However, because these groups of viewers are exposed to different numbers of ads, they should, according to our theory, differ in their familiarity with the shows.

Tables 2A–2D demonstrate that advertising improves the matching of consumers and products. Table 2A compares consumers who were not exposed to ads for any of the shows in a specific time slot with those who were exposed to at least one ad for each of the competing shows. The average gain in the match variable is 0.071 for the first group and 0.292 for the second.[20] The difference is significant at the 1% level. Thus, consumers who were exposed to ads are, indeed, more likely to choose shows that are suitable for them (than consumers who were not exposed to any ads).[21]

Whereas Table 2A focuses on consumers' choices among the shows of the three leading networks, Tables 2B–2D restrict the analysis to pairs of shows from these networks. The advantage

[18] The age match is slightly different. There are four age groups: teens, Generation-X, Baby Boomers, and older. The $Match_{i,j,t}$ variable gets a value of one when the age group of the individual and the show are the same. Otherwise, the index is equal to one minus one-half the number of age groups that separate the age group of the individual and the show's cast. Thus, for example, the match value of a teen watching a Generation-X show is 0.5.

[19] In other words, we compare viewers who were exposed to zero ads for each of the competing shows in a specific time slot with people who were exposed to one ad for each of those shows, and so forth.

[20] The gain is equal to the difference between the match of the consumer with the chosen show and the *average match* with the two competing shows. Even when we use the *maximum match* with the two competing shows (instead of the average match), we find similar results. Specifically, when we use the *maximum match*, the difference between the averages of the two groups is 0.252 and is different from zero at the 1% significance level. The time slots in the table are from the second part of the week (Wednesday–Friday). Time slots with news magazines are not included because these shows cannot be categorized based on the demographics of the cast.

[21] It is also worth noting that the average gain of the first group, 0.071, is different from zero at the 1% level. This suggests that, even without any exposure to advertising, consumers have some prior information about the shows.

TABLE 2 Advertising and Matching

(A) Effect of Exposure to Ads on the Gain in the "Match" between Viewers and Shows

	The Gain over the Average Match (across the Two Other Major Networks)			The Gain over the Highest Match that Can Be Reached in One of the Other Major Networks		
Number of Exposures to Ads for Each Network	Average	Standard Error	Number of Observations	Average	Standard Error	Number of Observations
0	0.071	0.019	1053	−0.226	0.022	1053
1 or more	0.292	0.033	517	0.026	0.039	517

(B) Effect of Exposure to Ads on the Gain in the "Match" between Viewers and Shows (ABC and CBS)

What Did the Individual Watch in the Previous Time Slot?	N_i^a	Average	Standard Deviation	Number of Obs.	*t* Statistics for the Difference with Respect to 0 Ads
Neither ABC nor CBS	0	0.086	0.749	451	—
	1	0.121	0.767	62	0.338
	2	0.136	0.839	11	0.196
ABC	0	0.255	0.743	475	—
	1	0.305	0.696	146	0.747
	2	0.375	0.762	32	0.864
CBS	0	−0.161	0.829	230	—
	1	0.216	0.826	51	2.947
	2	0.429	0.638	21	3.945

(C) Effect of Exposure to Ads on the Gain in the "Match" between Viewers and Shows (ABC and NBC)

What Did the Individual Watch in the Previous Time Slot?	N_i^a	Average	Standard Deviation	Number of Obs.	*t* Statistics for the Difference with Respect to 0 Ads
Neither ABC nor NBC	0	0.070	0.712	557	—
	1	0.207	0.722	87	1.649
	2	0.210	0.616	31	1.221
ABC	0	−0.210	0.743	362	—
	1	0.032	0.553	94	3.501
	2	0.017	0.590	29	1.952
NBC	0	0.045	0.506	441	—
	1	0.234	0.731	218	3.433
	2	0.061	0.603	99	0.245

(*Continued*)

of the latter comparison is that we can divide the consumers into three groups. The groups are based on the number of ads they were exposed to with respect to each of the competing shows: (i) no ads, (ii) exactly one ad, and (iii) more than one ad. In the previous comparison, the third group was too small for analysis (2%). Moreover, because (as we show later) lagged choices affect utility, we separately examine the cases that correspond to different lagged choices (i.e., not watching the networks, watching one, or watching the other). In all the cases, the average gain of the third group is, as expected, higher than the average gain of the first. Furthermore, in most cases, the average gain is the highest for the third group and the lowest for the first.

Tables 2A–2D bring initial evidence for the matching role of advertising,[22] but there are reasons to treat this evidence with caution. First, our measure of the match between consumers

[22] Additional nonstructural tests of informative advertising can be found in the working paper version of this study (Anand and Shachar, 2001). One such test exploits the fact that each television show spans multiple 15 minute time slots. We examine whether a viewer's tendency to switch away from a show after having watched it in the first 15 minutes decreases in ad exposures. The logic is that viewers who have been exposed to more ads are more informed, and thus

TABLE 2 Continued.

(D) Effect of Exposure to Ads on the Gain in the "Match" between Viewers and Shows (NBC and CBS)

What Did the Individual Watch in the Previous Time Slot?	N_i^a	Average	Standard Deviation	Number of Obs.	t Statistics for the Difference with Respect to 0 Ads
Neither CBS nor NBC	0	0.259	0.891	533	—
	1	0.259	0.892	85	0.000
	2	0.321	1.029	28	0.313
CBS	0	−0.211	0.870	266	—
	1	−0.034	0.913	89	1.602
	2	0.015	0.864	68	1.922
NBC	0	0.279	0.890	446	—
	1	0.466	1.002	208	2.301
	2	0.771	0.792	48	4.038

Notes:

(i) $Match_{i,j}$ is the demographic match between viewers and shows. This variable is based on three demographic characteristics: age, gender, and family status. It counts the number of characteristics that are identical for both the show and the individual. For example, for a Generation-X single female viewer and a Generation-X show with a single, male cast, $Match_{i,j} = 2$. For additional details, see footnotes 15 and 16 in the text.

(ii) The 1053 observations in the first row represent cases in which the individual was not exposed to any ad for any of the shows in a specific time slot. The 517 observations in the second row were exposed to at least one ad for *each* of the shows in that specific time slot.

(iii) The first and the fourth columns represent the average of the differences between the match with the show chosen by the individual ($\sum_{j=1}^{3} I\{C_i = j\} \cdot Match_{i,j}$) and a benchmark case. In the first column, the benchmark case is the average match across the two other networks ($0.5 \sum_{j=1}^{3} I\{C_i \neq j\} \cdot Match_{i,j}$). In the fourth column, the benchmark is the highest match that can be realized from one of the other networks ($Max_{j \neq C_i}(Match_{i,j})$). The second and fifth columns represent the standard errors of these averages.

In Tables 2B–2D: The first column (titled "Average") represents the average of the difference between the match of the individual with the show that she chose and her match with the show that she did not choose. The standard deviation of this difference appears in the second column, while the third column presents the number of observations in the row. The last column presents the t statistics for the hypothesis that exposure to one (or two) ad increases the differences between the match values chosen versus not.

and products is fairly crude. Second, we do not account here for unobserved preferences. For example, it is possible that some viewers do not watch a lot of TV and thus spend little time trying to get the right match, whereas others do watch a lot and try to get the right match. In such a case, the first type of viewers is less likely to watch ads and more likely to get a worse match, whereas the second will watch more ads and get a better match. In other words, in such a case, the relationship between exposure to ads and higher match values is not due to the matching role of ads. Such issues (i.e., crude measure of the match and unobservables) are resolved in the structural estimation. Specifically, the following section presents a model that identifies additional implications of the effect of advertising through the information set, includes a more precise and flexible measure of the match, and allows us to test this theory directly by structural estimation.

3. The model

■ This section introduces the utility function, the information set, and the implications of the model.

We study differentiated products and heterogeneous consumers in a setting that is quite similar to Berry, Levinsohn, and Pakes (1995; hereafter BLP). Following Lancaster (1971), we formulate consumer utility over products as a function of individual characteristics and the

less likely to be disappointed and switch away after watching the beginning of the show; in other words, informative advertising reduces regret. The data strongly support this hypothesis. Furthermore, even when we control for the fraction of time that individuals watch TV (because this "personal taste for TV" variable may be correlated with both advertising exposure and switching decisions), we still find strong support for our hypothesis.

attributes of those products. Our discrete-choice model has a random utility as in McFadden (1981). Unlike BLP, individuals are uncertain about product attributes and, as in Grossman and Shapiro (1984), advertising is informative. Another difference between our setting and that in BLP arises from our use of individual-level panel data. This both facilitates a richer treatment of heterogeneity and requires that the model be extended to account for its dynamic aspects.

□ **The setup.** Let $j = 0, \ldots, J$ index the alternatives of the individuals, where $j > 0$ index the competing multiproduct firms (i.e., the television networks ABC, CBS, NBC, and Fox), and $j = 0$ is the "outside good" (i.e., not watching any of the network shows). In each period t, which is also called a "time slot" and lasts 15 minutes, each of the J networks offers a single product (i.e., a television show). Each show is offered only once within the studied time frame. In other words, the firms are multiproduct across time, but in any t they are single-product firms. To avoid excessive and cumbersome notation, we do not use a show-specific index. Note, however, that whereas j is indexing a network, the combination of j, t (i.e., a specific time slot on network j) represents a specific show.

There are I individuals who are indexed by i. In each period t, individual i makes a choice. The expression $y_{i,t} = j$ is our notation for the event that the choice of individual i at time t is alternative j.

□ **The utility.**

Utility from a TV show. The utility from a TV show is a function of: (i) the match between the consumer's preferences and show attributes; (ii) the direct effect of advertising; and (iii) aspects related to the dynamic nature of the model, such as state dependence. Equation (1) describes the general structure of the utility.

For the viewing alternatives ($j > 0$), the utility of individual i in period t (where any combination of j and t defines a show) is

$$U_{i,j,t} = x_{j,t}\beta_i + (\xi_{j,t} + \varepsilon_{i,j,t}) + g_i\left(N^a_{i,j,t}\right) + h_{i,j,t}(y_{i,t-1}) + \upsilon_{i,j}. \tag{1}$$

Viewer-show match. The row vector $x_{j,t}$ captures observed show attributes, and the parameter vector β_i stands for the preferences of the viewer. As we clarify later, in this model the viewer is uncertain about $x_{j,t}$ (which is observed to the researcher but not the viewer) but is not uncertain about her preferences β_i, which she fully knows. The particular formulation of the interaction $x_{j,t}\beta_i$ here is

$$\begin{aligned} &\beta_{Gender} Gender_{i,j,t} + \sum_{k=0}^{2} \beta_{Age,k} Age_{k,i,j,t} + \beta_{Family} Family_{i,j,t} + \beta_{RaceIncome} RaceIncome_{i,j,t} \\ &\quad + \sum_{Genre=1}^{5} x^{Genre}_{j,t}\left(\beta_{Genre} z^0_i + \upsilon^{Genre}_i\right). \end{aligned} \tag{2}$$

The first line represents the effect of cast demographics on choices. Each variable in this line captures a match between the demographics of the show's cast and the viewer. All these variables are binary, with zero-one values. Specifically, the variable $Gender_{i,j,t}$ equals one if the gender of viewer i and the cast of show j, t is the same; $Age_{o,i,j,t}$ equals one if the age group of viewer i and the cast of show j, t is the same[23]; $Age_{1,i,j,t}$ equals one if the distance between the age group of viewer i and the cast of show j, t is one; $Age_{2,i,j,t}$ is defined accordingly; and $Family_{i,j,t}$ equals one if viewer i lives with her family and show j, t is about family matters. The race of the viewer is not

[23] The age groups are (i) younger than 18 years old, (ii) between 18 and 34 years old, (iii) between 35 and 49 years old, and (iv) older than 49 years old.

included in our data set, and we approximate it with her income.[24] Thus, the race match variable, $RaceIncome_{i,j,t}$, is equal to the interaction between $Income_i$ and a binary variable that equals one if one of the main characters in show j, t is African-American.

Previous studies have demonstrated that viewers have a higher utility from shows whose cast demographics are similar to their own. Thus, one should expect to find that (i) $\beta_{Age0} > \beta_{Age1} > \beta_{Age2}$, (ii) $\beta_{Gender} > 0$, (iii) $\beta_{Family} > 0$, and (iv) $\beta_{RaceIncome} < 0$.

The second line represents the effect of show genre on choices. The show genres included in $x_{j,t}^{Genre}$ are defined and described in Section 2. The taste parameter is a function of observed and unobserved individual characteristics, z_i^0 and υ_i^{Genre}, respectively. For notational simplicity, we hereafter let υ_i include any individual-specific unobserved parameter (including υ_i^{Genre}). The observed variables included in the eight-dimensional vector z_i^0, defined in Table 1, are the individual's age, gender, income, education, and family status. This vector multiplies the taste parameter β_{Genre} (which is an eight-dimensional row vector). Thus, each interaction between show genre and individual characteristics is captured through a unique parameter. For example, the interaction between an action drama show and a female viewer is captured via β_{AD}^{Female}. All the other parameters are denoted accordingly.[25]

Because some of the show attributes are unobserved by the researcher, some components of the match element are unobserved as well. The parameter $\xi_{j,t}$ can be thought of as the mean (across individuals) of these unobserved matches, and $\varepsilon_{i,j,t}$ can be thought of as the personal deviation from that mean. Both the observed and the unobserved product attributes, $x_{j,t}$ and $\xi_{j,t}$, respectively, differ across shows, but are constant for the duration of each show.

The direct effect of advertising. $N_{i,j,t}^a$ denotes the number of exposures to advertisements for show j, t by individual i. The individual-specific function $g_i(\cdot)$ represents the direct effect of ad exposures on the utility. This is the modelling approach adopted by previous empirical studies and often termed the "persuasive" effect (see, e.g., GS). Despite the inclusion of the direct effect of advertising on utility in many empirical studies in economics, most justifications for it come from the psychological and marketing literature.[26] Specifically, Zajonc (1968) and studies that followed it demonstrate, largely through experiments, that advertising repetition may lead to a preference for the advertised product even if consumers do not absorb information on product benefits. Separately, Krugman (1968) argues that repeated ad exposure creates familiarity with a product, which in turn leads to a subsequent liking for that product.

Even though our main objective is to demonstrate the effect of advertising through the information set, the model includes this direct effect on the utility. Besides the behavioral justification, there are other salient reasons for including such an effect. First, it enables a comparison between previous works and ours. Second, excluding the direct effect of ads on the utility would force ads to affect choices in the model only through the information set. Thus, any evidence about the informative role of advertising might be suspected as resulting from misspecification.

[24] The proportion of African-Americans in the highest income category is disproportionately low, while it is disproportionately high in the lowest income category. This relationship persists for all income categories in between as well (U.S. Bureau of the Census, 1995). Nielsen designed the sample to reflect the demographic composition of viewers nationwide and used 1990 Census data to achieve the desired result. We found that the income categories and the proportion of African-Americans in the Nielsen data closely match those in the U.S. population (National Reference Supplement, 1995). Although our data set does not include information about race, Nielsen has it and reports its aggregate levels.

[25] Notice that an alternative way to denote the consumer-product match in (1) is $x'_{i,j,t}\beta + x''_{j,t}\upsilon''_i$. Such notation makes explicit that many of our variables are individual-specific covariates such as the interaction between an action drama show and a female viewer. However, we opt to denote the match as $x_{j,t}\beta_i$ because it is more concise.

[26] An exception is Becker and Murphy (1993), who suggest that advertising is a complement to the product being advertised, in which case one can justify including advertising directly in the utility function. But because most advertisements for television shows are excerpts from the show itself, the Becker-Murphy reasoning probably has limited relevance to the television application here.

The specific functional form of $g_i(\cdot)$ used in the empirical analysis is quadratic:

$$g_i\left(N^a_{i,j,t}\right) = \rho_{i,1,t} N^a_{i,j,t} + \rho_{i,2,t}\left(N^a_{i,j,t}\right)^2, \tag{3}$$

with $\rho_{i,2,.} < 0$ corresponding to the often-termed "wear-out" effect of advertisements.

The parameter $\rho_{i,1,t}$ is equal to $\rho_{i,1,MT} MT_t + \rho_{i,1,WF} WF_t$, where MT_t and WF_t are binary variables equal to one for shows that aired on Monday–Tuesday and Wednesday–Friday, respectively. The parameter $\rho_{i,2,t}$ is defined accordingly. We allow the advertising parameters to differ across these two parts of the week to account for the problem of missing data mentioned in Section 2. The ρ parameters are allowed to vary across consumers for observed and unobserved reasons. For example, $\rho_{i,1,MT} = \rho_1 z^0_i + \upsilon^\rho_{i,1,MT}$. Notice that although the $g_i(\cdot)$ function is not central to our theory, its formulation here is richer than in most prior work.

State dependence. The last two elements in the utility represent state dependence and a related unobserved heterogeneity parameter. These elements capture important dynamic aspects of the empirical example. However, they are not important for our general model of advertising because the main implications of the model can be illustrated even in a static setting.[27]

Previous studies of television-viewing choices find strong evidence of state dependence *between shows on the same network*.[28] Indeed, our data reveal that, on average, 65% of viewers who were watching a show on network j watched the *next* show on the same network. State dependence is obviously not the only explanation for this finding. For example, the tendency by the networks to schedule similar shows in sequential time slots might also lead to the high persistence rate in choices.[29] However, it turns out that controlling for observed and unobserved show attributes does not eliminate the support for state dependence (Goettler and Shachar, 2001). Previous studies suggested several explanations for this state dependence. For example, one distinctive feature of watching television (versus other leisure activities, such as sports or social events) is its passive nature. Indeed, for many people, watching television is a way to relax and, thus, actively flipping channels might be annoying.

We formulate the state-dependence function, $h_{i,j,t}(y_{i,t-1})$, as[30]

$$\begin{gathered}(\delta_i + \delta_{j,t}) I\{y_{i,t-1} = j\} + (\delta_{InProgress} Continuation_{j,t}) I\{y_{i,t-1} \neq j\},\\ \text{where } \delta_i = z^\delta_i \delta^z + \upsilon^\delta_i\\ \text{and } \delta_{j,t} = \left(\delta_{Cont} + x^{Genre}_{j,t}\delta^x\right) Continuation_{j,t} + \delta_{First15} First15_{j,t} + \delta_{Last15} Last15_{j,t}.\end{gathered} \tag{4}$$

We allow the state dependence to vary across individuals, δ_i, and products, $\delta_{j,t}$. The variation across individuals comes from observed and unobserved sources, z^δ_i and υ^δ_i, respectively. The vector z^δ_i includes all the variables in z^0_i and the viewer's cable subscription status ($Basic_i$ and $Premium_i$).[31] For example, a common perception is that men switch channels more frequently than women. The structure above allows for such differences across individuals.

One might expect that the persistence during a show depends on the type of program. For example, it is likely that such persistence is higher during shows with a plot (such as romantic dramas) than during other shows (such as news magazines). The persistence during a show might also depend on the time that the individual has already spent watching it. For example, a viewer

[27] See our working paper (Anand and Shachar, 2001).

[28] See, for example, Rust and Alpert (1984).

[29] This strategy, termed "homogeneity," is mostly followed by the networks from 8:00 to 10:00 p.m. However, there are frequent deviations from this strategy. Furthermore, in most cases, the shows that start at 10:00 p.m. are dissimilar to those that preceded them.

[30] The variables $Continuation_{j,t}$, $First15_{j,t}$, and $Last15_{j,t}$ are binary variables that get the values 0 and 1. They are equal to 1 if the following conditions hold: for $Continuation_{j,t}$ if the show on j started at least 15 minutes ago; for $First15_{j,t}$ if the show on j started in the previous 15 minute time slot; and for $Last15_{j,t}$ if the show on j is at least one hour long and will end within 15 minutes.

[31] The binary variable $Basic_i$ is equal to one for the one third of the population that has access only to basic cable offerings, and the binary variable $Premium_i$ is equal to one for the one third of the population that has both basic and premium cable offerings.

who watched 45 minutes of a one hour drama is unlikely to switch away at this point. These hypotheses are integrated into the $h(\cdot)$ function via the third line of equation (4). Finally, the cost incurred by a viewer who tuned in to a program after it already started is $\delta_{InProgress}$.

The state dependence is a function of an endogenous variable—the previous choice. If this choice depends on unobserved variables (or parameters) that also affect the current decision, the estimates of the state dependence would be inconsistent. A standard approach to dealing with this problem (in structural models) is to model these temporally persistent unobservables and integrate them out from the history probability. Thus, we include in the utility (1) a network-individual unobserved match parameter, $\upsilon_{i,j}$. This parameter does not have an index t and, thus, for each individual, is common to all the shows offered by network j.[32]

Utility from the outside alternative. The utility from the outside alternative is

$$U_{i,0,t} = (z_{i,t}\gamma + \upsilon_{i,0,Hour(t)}) + (\xi_{0,t} + \varepsilon_{i,0,t}) \\ + (z_i^\delta \delta^z + \upsilon_i^\delta + \delta_{0,t}) I\{y_{i,t-1} = 0\}, \tag{5}$$

where $z_{i,t}$ includes all the variables in z_i^0 and the variables $Basic_i$, $Premium_i$, All_i, and $Same_{i,t}$. The cable subscription status is included because the outside alternative includes the option of watching nonnetwork shows—viewers with basic or premium cable have a larger variety of choices, which can lead to a higher utility. The variable All_i (and $Same_{i,t}$) is equal to the average time that the individual watched television (and in the corresponding time slot t) during the previous days of the week. Individuals' tendencies to watch television cannot be fully explained by their demographic characteristics. Thus, their prior viewing habits (All_i and $Same_{i,t}$) and the personal unobserved parameters $\upsilon_{i,0,Hour(t)}$ are designed to capture other sources of such differences. Specifically, we allow $\upsilon_{i,0,Hour(t)}$ to differ across the three hours of each night. The $(\xi_{0,t} + \varepsilon_{i,0,t})$ terms are analogous to the ones defined above for the J network alternatives.[33]

The personal state-dependence effect $(z_i^\delta \delta^z + \upsilon_i^\delta)$ enters (5) exactly the same as in (4) because it is meant to represent behavioral attributes intrinsic to individuals. For example, some people tend to be more restless or more active and, thus, intrinsically like to switch between activities. Such behavior would not be specific to switching between television shows only. However, we add to this personal effect a parameter that is unique to the outside alternative—$\delta_{0,t}$. Furthermore, because the outside alternative includes the option to watch nonnetwork shows, many of which end on the hour, we allow this parameter to change "on the hour." Specifically, we set $\delta_{0,t}$ to be equal to $\delta_0 + \delta_{Hour} NineTen_t$, where the binary variable $NineTen_t$ equals one for t that is either 9 p.m. or 10 p.m.

□ **Information set.** Unlike most discrete-choice models, we assume that the individual is uncertain about product attributes, $\xi_{j,t}$ and $x_{j,t}$, and, thus, about $(\xi_{j,t} + x_{j,t}\beta_i)$. We denote this expression by $u_{i,j,t}^{att}$ and term it "attribute utility," because it represents the contribution of product attributes to utility. Specifically,

$$u_{i,j,t}^{att} \equiv \xi_{j,t} + x_{j,t}\beta_i. \tag{6}$$

[32] Whereas the state dependence between shows might be surprising, the persistence of choices during a show is not. Straightforward explanations are that the show attributes (observed or unobserved) remain the same, and that viewers get hooked on a plot as the show progresses. This and similar hypotheses are integrated into the specific functional form as described in equation (4). Furthermore, our model accounts not only for an unobserved individual-*network* parameter but also for an unobserved individual-*show* parameter, as explained later in Section 3.

[33] Instead of estimating $\xi_{0,t}$ for each of the 60 time slots, we impose the following restriction:

$$\xi_{0,t} = \xi_{0,t+12} = \xi_{0,t+24} = \xi_{0,t+36} = \xi_{0,t+48} \text{ for } t = 1, \ldots, 12.$$

This implies that the outside utility for the time slot between 8:00 and 8:15, for example, is the same across all the nights of the week. This allows us to identify the expected increase in the outside utility during the night with 48 fewer parameters.

The information set includes: (i) a prior distribution of products' attributes; and (ii) product-specific signals such as advertising and word of mouth. In other words (and as we elaborate below), consumers have two types of nonadvertising information: product-specific signals and firm-specific information (that determines the prior distribution).

Prior distribution. Even prior to getting any product-specific signal, an individual has some knowledge about the distribution of $\xi_{j,t}$ and $x_{j,t}$. In general, one might assume that the prior distribution of $u^{att}_{i,j,t}$ is[34]

$$u^{att}_{i,j,t} \sim N\left(\mu_{i,j}, \frac{1}{\varsigma^{\mu}_{i,j}}\right), \tag{7}$$

where, by definition, $\mu_{i,j} = E_t(\xi_{j,t}) + E_t(x_{j,t})\beta_i$.[35] This means that although the individual is uncertain about $u^{att}_{i,j,t}$, she knows the expected value and the variance of $\xi_{j,t}$ and $x_{j,t}$ for each multiproduct firm. Indeed, TV networks, like other multiproduct firms such as automakers, are known to have distinct profiles. Thus, it is reasonable to assume that consumers, although uncertain about the attributes of each product, know the firms' profiles.

As demonstrated later, a consequence of this assumption is that consumers rely on multiproduct firms' profiles when forming their expectations about specific products. Such behavior seems quite reasonable in many contexts. For example, a consumer thinking of purchasing a Corolla would use her information about the Camry (or other Toyota cars) to update her expectations. In the TV context as well, it appears that this approach is quite sensible. Indeed, the network TV industry serves Mankiw (1998) in his *Economics* textbook as a good example of the informational role of multiproduct firms. Referring to multiproduct firms as "brands," he writes: "Establishing a brand name—and ensuring that it conveys the right information—is an important strategy for many businesses, including TV networks."[36]

Beyond its economic plausibility, allowing the prior to depend on the multiproduct firms' profiles also has an econometric justification. Specifically, in prior structural estimation (Anand and Shachar, 2004), we show that this assumption is supported by the data on television-viewing choices. Later in the structural estimation, we return to a robustness check of this assumption—in Section 5, we relax it and demonstrate that the empirical results are not sensitive to it.

The assumption that the prior distributions depend on the multiproduct firms' profiles is implemented as follows. In the estimation, we set the moments of the distribution of product attributes to be equal to those of the empirical distribution. Furthermore, we restrict the prior distribution to account for a known strategy employed by the networks. Specifically, shows aired by the television networks between 10:00 and 11:00 p.m. tend to be dissimilar to those aired between 8:00 and 10:00 p.m. For example, sitcoms are not broadcast after 10:00 p.m. on any night. Because this strategy is well known, viewers are likely to have different prior beliefs about the scheduling for these two parts of the night. We account for that by allowing the prior distribution to differ not only across the networks but also across the different parts of the night.

Specifically, the prior distribution for each part of the night depends only on the distribution of the attributes of the shows that are broadcast during that part. That is, for example, for shows aired between 8:00 and 10:00 p.m., $\mu^{8-10}_{i,j} = \frac{1}{40}\sum_{t\epsilon t^{8-10}} u^{att}_{i,j,t}$, where t^{8-10} is the set of all the time slots between 8:00 and 10:00 p.m. during the week.

[34] Although the normality assumption is made mostly for convenience, it seems reasonable because it relates to the attribute utility, which is a weighted average of various variables. We should also point out that in the application, we base the moments of the distribution on the data.

[35] Notice that the expectation $\mu_{i,j}$ and the precision $\varsigma^{\mu}_{i,j}$ differ across individuals because the taste parameter β_i is individual specific.

[36] A *New York Times* article (September 20, 1996) that he cites reads: "In television, an intrinsic part of branding is selecting shows that seem related and might appeal to a certain audience segment. It means 'developing an overall packaging of the network to build a relationship with viewers, so they will come to expect certain things from us,' said Alan Cohen, executive vice-president for the ABC-TV unit of the Walt Disney Company in New York."

Product-specific signals. The individual receives product-specific signals on product attributes from various sources, such as word of mouth, previous experience with the product, media coverage, and advertising. In order to focus on the informational role of advertising, we separate the advertising signals from the miscellaneous ones.

Miscellaneous signals. The individual receives $N^m_{i,j,t}$ miscellaneous product-specific signals. These signals are i.i.d. Specifically, each signal is distributed as

$$\tilde{S}^m_{i,j,t,n} = u^{att}_{i,j,t} + \tilde{\omega}^m_{i,j,t,n} \text{ for any } 1 \leq n \leq N^m_{i,j,t}, \text{ and where } \tilde{\omega}^m_{i,j,t,n} \sim N\left(0, \frac{1}{\varsigma^m}\right). \tag{8}$$

We assume that these signals are noisy ($\frac{1}{\varsigma^m} > 0$) and unbiased. The noisiness can result from various sources. For example, neither media coverage nor word of mouth is a very precise source of information.[37]

Our data do not include $N^m_{i,j,t}$.[38] Thus, we can, without loss of generality, rewrite (8) as

$$\tilde{S}^m_{i,j,t} = u^{att}_{i,j,t} + \tilde{\omega}^m_{i,j,t} \text{ where } \tilde{\omega}^m_{i,j,t} \sim N\left(0, \frac{1}{\varsigma^m_{i,j,t}}\right), \tag{9}$$

where $\varsigma^m_{i,j,t} \equiv N^m_{i,j,t}\varsigma^m$.

The presence of miscellaneous signals implies that even without any ad exposures, consumers have product-specific information. Furthermore, this familiarity with products is heterogeneous across consumers and shows (as embodied in the differences in the number of miscellaneous signals $N^m_{i,j,t}$). We expect, for example, that the $\varsigma^m_{i,j,t}$ of a show that has been aired for many years is higher than that of a new program, and that some consumers are more knowledgeable than others. Thus, we allow $\varsigma^m_{i,j,t}$ to be a function of shows' characteristics, as well as an individual-specific unobserved parameter, $\upsilon^m_{i,j}$. Specifically, we formulate $\varsigma^m_{i,j,t}$ as $\upsilon^m_{i,j} + \varsigma^m_{New} New_{j,t} + \varsigma^m_{Veteran} Veteran_{j,t}$, where the binary variables $New_{j,t}$ equals one for shows that are in their first season and were not rated in the top 20 shows during any of the previous three weeks, and $Veteran_{j,t}$ equals one for shows in at least their fifth season (that remained in the same time slot as in the previous season) or were rated in the top 20 during each of the previous three weeks.

Furthermore, the familiarity of an individual with a show is also likely to be correlated with her preferences. For example, an individual who likes NBC is likely to be quite familiar with its shows. In the estimation, we account for the correlation between familiarity and tastes to result from both observable and unobservable sources. Specifically, we allow for a correlation between $\upsilon^m_{i,j}$ and the other elements in υ_i.

Dynamic learning through previous experience with a product is the focus of various studies (Eckstein, Horsky, and Raban, 1988 and Crawford and Shum, 2005, for example). Those studies rely on consumer choice data that span multiple weeks and multiple purchase occasions. Unlike these studies, we have only one week of data, and each show is offered only once during this week. Thus, previous experience is unobserved and is incorporated into the model through the miscellaneous signals.[39]

To demonstrate the combined impact of the prior distribution and the miscellaneous signals, consider a consumer who was not yet exposed to any ad. One can think of such a consumer forming expectations about the attributes of a show, say on CBS, in two steps. Her "initial prior" about the show depends on what attributes a CBS show *typically* has. This is the information contained in the network profile. Second, she receives noisy information from miscellaneous

[37] Indeed, as discussed in Section 2, it is reasonable to assume, for the television application, that even past experience is a noisy signal.

[38] These unobserved miscellaneous signals also include, among other things, elements such as word-of-mouth and exposure to advertisements that are not in the data set.

[39] Learning through word of mouth has been studied by Mahajan, Muller, and Wind (2000) and Ching (2000), and learning via media coverage by Bond and Kirshenbaum (1998).

signals. The resulting posterior distribution can be thought of as her "preadvertising prior." Notice that her pre-advertising prior for any given show includes information both on attributes of that show (through the miscellaneous signals) as well as on attributes of other shows on that network (through the network profile).

Advertising signals. The content of each advertisement serves the individual as a signal on product attributes. These signals are i.i.d. Specifically, each signal is distributed as

$$\tilde{S}^a_{i,j,t,n} = u^{att}_{i,j,t} + \tilde{\omega}^a_{i,j,t,n} \text{ for any } 1 \leq n \leq N^a_{i,j,t}\text{, and where } \tilde{\omega}^a_{i,j,t,n} \sim N\left(0, \frac{1}{\varsigma^a}\right). \tag{10}$$

We assume that the signals are noisy (that is, $\frac{1}{\varsigma^a} > 0$) and unbiased.[40] The noisiness of advertising is well accepted and documented in Jacoby and Hoyer (1982, 1989). We assume that the signals are independent for two reasons: (i) in some cases, a show might be promoted via different advertisements (i.e., two ads for the same show might have different content); and (ii) different exposures to the same advertisement can lead to different impressions. The independence assumption does not affect our qualitative results.

The effect of advertisements through the information set is captured by ς^a. If $\varsigma^a = 0$, then advertisements are too noisy to convey any information about product attributes. In other words, when $\varsigma^a = 0$, the information sets of two individuals who differ only in N^a are the same. However, when $\varsigma^a > 0$, the information sets of such consumers differ. Thus, ς^a is the key parameter of interest in the empirical study.

□ **Expected "attribute utility" and implications.** The expected "attribute utility" is the mean of the posterior probability. It is denoted by $\mu^p_{i,j,t}$ and it is equal to (DeGroot, 1989)

$$\mu^p_{i,j,t} = \frac{1}{\varsigma^p_{i,j,t}}\left[\varsigma^\mu_{i,j}\mu_{i,j} + \varsigma^m_{i,j,t}S^m_{i,j,t} + \varsigma^a\sum_{n=1}^{N^a_{i,j,t}} S^a_{i,j,t,n}\right], \tag{11}$$

where $\varsigma^p_{i,j,t} = \varsigma^\mu_{i,j} + \varsigma^m_{i,j,t} + N^a_{i,j,t}\varsigma^a$, and $S^a_{i,j,t,n}$ and $S^m_{i,j,t}$ are the realizations of the signals.[41] This means that the expected attribute utility is a weighted average of the three sources of information: the networks' profiles, the miscellaneous signals, and the ads. The weight placed on each source of information is equal to its precision (compared to the precision of all information sources together).

This model assumes that consumers are uncertain about product attributes, and that advertising is an element in their information set. The rest of this section traces out the behavioral implications of these two assumptions. These implications allow the researcher to examine each of these model assumptions in the data.

In order to address these questions and derive the behavioral implications, one needs to distinguish between the knowledge of the individual and that of the researcher. Whereas the individual observes the realizations of the signals, but not $u^{att}_{i,j,t}$, the researcher does not observe the signals but has an estimate of $u^{att}_{i,j,t}$ (because he has estimates of $\xi_{j,t}$ and β_i, and the data consist of $x_{j,t}$).

It is easy to show that we can rewrite (11) from the researcher's point of view as

$$\begin{aligned} &\mu^p_{i,j,t} = \left[\mu_{i,j} + \lambda_{i,j,t}\left(u^{att}_{i,j,t} - \mu_{i,j}\right)\right] + \sigma^\omega_{i,j,t}\omega_{i,j,t}, \\ &\text{where } \lambda_{i,j,t} \equiv \frac{\varsigma^m_{i,j,t} + N^a_{i,j,t}\varsigma^a}{\varsigma^p_{i,j,t}},\ \sigma^\omega_{i,j,t} \equiv \frac{\sqrt{\varsigma^m_{i,j,t} + \varsigma^a N^a_{i,j,t}}}{\varsigma^p_{i,j,t}},\ \text{and } \omega_{i,j,t} \sim N(0, 1). \end{aligned} \tag{12}$$

[40] Rational consumers know the distribution from which a signal is drawn, so assuming the signals are unbiased is a location normalization.

[41] Notice that $\frac{1}{\varsigma^p_{i,j,t}}$ is the variance of her posterior distribution.

When the individual is fully informed (for example, when $(\varsigma^m)^{-1} = 0$), $\lambda_{i,j,t} = 1$ and $\sigma^{\omega}_{i,j,t} = 0$. In this case the expected attribute utility, $\mu^p_{i,j,t}$, is equal to the actual attribute utility, $u^{att}_{i,j,t}$. In general, $\lambda_{i,j,t}$ can be thought of as a measure of how well informed the individual is.

Consumer uncertainty. Now, it can be seen from (12) that the profile of the TV network, $\mu_{i,j}$, affects the viewer's choices if and only if $\lambda_{i,j,t} < 1$. Thus, data that reveal the viewing probability to be a function of $\mu_{i,j}$ would imply that the viewer is uncertain about show attributes. Indeed, in Anand and Shachar (2004) we have employed this identification strategy to show that TV viewers are not fully informed.

Informative advertising. In this subsection, we trace out certain implications of informative advertising that allow us to estimate the relevant parameters in the data. Each of these implications is unique to the informative role of advertising and behaviorally distinct from the direct effect of advertising on utility, allowing separation of the two effects empirically. The discussion here is closely related to the discussion on identification in Section 4.

First, notice that when advertising is informative (i.e., $\varsigma^a > 0$), $\lambda_{i,j,t}$ (the measure of how well informed the individual is) increases in the number of ad exposures, as expected.

In order to examine the additional behavioral implications of informative advertising, it is useful to focus on the following derivative:

$$\frac{\partial \mu^p_{i,j,t}}{\partial N^a_{i,j,t}} = \left[\frac{\varsigma^a}{\left(\varsigma^p_{i,j,t}\right)^2}(1 - \lambda_{i,j,t})\right]\left(u^{att}_{i,j,t} - \mu_{i,j}\right) + \frac{\partial \sigma^{\omega}_{i,j,t}}{\partial N^a_{i,j,t}}\omega_{i,j,t}. \tag{13}$$

We focus our attention on the first element in equation (13) because, as explained below, the second one is somewhat similar to Erdem and Keane (1996). Notice that whereas the first element is observed by the researcher, the second is not.

Consumption deterrence. The first implication is that in some cases, exposure to advertising should decrease the consumer's tendency to purchase the promoted product. Notice that the sign of the effect of advertising through the information set is determined by $(u^{att}_{i,j,t} - \mu_{i,j})$ because $[\frac{\varsigma^a}{(\varsigma^p_{i,j,t})^2}(1 - \lambda_{i,j,t})]$ is always positive.[42] Because this sign might be either positive or negative, this implies that in some cases, exposure to advertising decreases the purchase probability. Specifically, when $u^{att}_{i,j,t} < \mu_{i,j}$, the purchase probability is a decreasing function of $N^a_{i,j,t}$. In other words, advertising can deter consumption. This (first) behavioral implication of this model is novel even with respect to models that include advertising in the information set.

The intuition behind this result is simple. Whenever the match between a consumer and a product is (relatively) low, any product-specific information will decrease the consumer's tendency to buy the product. Advertising provides such information.

Matching. The second implication, which immediately follows from the first, is that informative advertising improves the match between products and consumers. Indeed, it is easy to show that advertisement about *any* product improves the matching process. Again, the intuition is straightforward. By reducing consumers' tendencies to purchase products that do not fit their preferences well and increasing their tendency to buy those that do, advertising increases the consumer's utility on average.

Although this second implication was explored empirically in Section 2, the first implication is examined later, using the structural estimates.

Advertising effectiveness. The effect of ads through the information set is quite different from the direct effect via the utility. In addition to the implications regarding consumption deterrence and matching, the first element of (13) embodies several implications about the effectiveness of informative advertisements. For example, informative ads are especially effective (i) for products

[42] Notice that if the individual is fully informed, $\lambda_{i,j,t} = 1$, and this element is equal to zero.

whose attributes are very different from the average product of the firm, and (ii) for lesser-known products.

For products whose attributes are very different from the average product of the firm, $|u^{att}_{i,j,t} - \mu_{i,j}|$ is large. In such a case, (13) implies that a change in $N^a_{i,j,t}$ leads to a large change in the expected attribute utility. For lesser-known products, $\varsigma^m_{i,j,t}$ is small. It is easy to show that in that case $[\frac{\varsigma^a}{(\varsigma^p_{i,j,t})^2}(1 - \lambda_{i,j,t})]$ is large and, thus, the effectiveness of ads is large. The intuition behind this result is that miscellaneous signals and advertising signals can be thought of as informational substitutes because they are both product specific. Conversely, when $\varsigma^m_{i,j,t}$ is larger, the added value of the informative advertising signal is smaller.[43]

Finally, notice that ad effectiveness is also a function of the diversity of products offered by the multiproduct firm and, obviously, of the precision of the ads, $\varsigma^\mu_{i,j}$ and ς^a, respectively.

Thus far, we have discussed the first element in (13). The second element, $\frac{\partial \sigma^\omega_{i,j,t}}{\partial N^a_{i,j,t}}\omega_{i,j,t}$, is unobserved and, thus, its behavioral implications are limited in our setting. It implies that informative advertising reduces the variation in consumer choice probabilities conditional on the observable variables. Although the derivative of $\sigma^\omega_{i,j,t}\omega_{i,j,t}$ with respect to $N^a_{i,j,t}$ provides little insight into the role of informative advertising, the inclusion of $\sigma^\omega_{i,j,t}\omega_{i,j,t}$ in the model is quite important for other reasons. State dependence *during* a show might result from an individual-*show*-specific unobserved variable. The element $\sigma^\omega_{i,j,t}\omega_{i,j,t}$ is exactly such an unobservable. Notice that although both $\sigma^\omega_{i,j,t}$ and $\omega_{i,j,t}$ have an index t, they do not vary during a show.

Informative versus persuasive advertising. The discussion above makes clear that the behavioral implications of the effect of advertising through the information set are very different from the effect via the utility. The expected utility provides a slightly more formal way to examine the differences between the two avenues through which ads affect choices. The expected utility of the individual, denoted by $u_{i,j,t}$, is

$$u_{i,j,t} = \left[\mu_{i,j} + \lambda_{i,j,t}\left(u^{att}_{i,j,t} - \mu_{i,j}\right)\right] + \sigma^\omega_{i,j,t}\omega_{i,j,t} + g_i\left(N^a_{i,j,t}\right) + h_{i,j,t}(y_{i,t-1}) + \upsilon_{i,j} + \varepsilon_{i,j,t} \text{ for } j > 0. \tag{14}$$

Thus, the derivative of the expected utility with respect to the number of ad exposures is

$$\frac{\partial u_{i,j,t}}{\partial N^a_{i,j,t}} = \left[\frac{\varsigma^a}{\left(\varsigma^p_{i,j,t}\right)^2}(1 - \lambda_{i,j,t})\right]\left(u^{att}_{i,j,t} - \mu_{i,j}\right) + \frac{\partial \sigma^\omega_{i,j,t}}{\partial N^a_{i,j,t}}\omega_{i,j,t} + \frac{\partial g_i\left(N^a_{i,j,t}\right)}{\partial N^a_{i,j,t}}. \tag{15}$$

The third element in equation (15) represents the direct effect of advertising on the utility. This is the standard role of (persuasive) advertising in previous models. Notice that its formulation is somewhat enriched here, to allow advertising effectiveness to vary across consumers for observed and unobserved reasons.

The distinction between the direct effect of ads through the utility and its effect through the information set is evident from (15). Related to this, the previous subsection discussed how Bayesian learning introduces certain behavioral implications and variables that are unique to the effect of ads through the information set. These include: (i) the "consumption deterrence" effect of informative advertising, and predictions on when this should be observed; (ii) the dependence of ad effectiveness on multiproduct firm profiles through $|u^{att}_{i,j,t} - \mu_{i,j}|$; and (iii) the impact on ad effectiveness of other elements in the information set (for example, consumers' familiarity with products and the diversity of products offered by the multiproduct firm). Each of these behavioral implications is unique to the informative effect of advertising and enables one to distinguish it from the direct effect of ads on the utility.

This also clarifies that the separation between the effect of ads through the information set and its direct effect through the utility does not come from functional-form restrictions. Indeed, because the heterogeneity in the direct effect is not correlated with the match parameters, even if

[43] In Shachar and Anand (1998), we have examined this implication and found that ad effectiveness is, indeed, larger for lesser-known products.

the formulation of the $g_i(\cdot)$ function in (15) were enriched further (for instance, by allowing it to differ across programs), the separation between the two effects can still be obtained.

□ **Forward looking.** In the discussion above, we assumed that the individual is myopic and, thus, maximizes her per-period expected utility. A forward-looking individual, who experiences state dependence, considers the future consequences of her current choices. Although forward-looking behavior is important in many decision-making contexts such as portfolio investment and job search, it seems less critical in our setting. Although some individuals may plan their viewing for the entire night accounting for the consequence of state dependence in later periods, we tend to believe, like previous studies, that such forward-looking viewers are rare.[44]

4. Estimation and identification issues

■ This section consists of four subsections. The first discusses the endogeneity problem in advertising models and describes our solution to this problem. The likelihood function is constructed in the second subsection, and our simulation approach follows. The last subsection discusses the identification of the model's parameters.

□ **Endogeneity.** Advertising spending and targeting are chosen strategically by firms. This introduces two potential sources of endogeneity in our data. First, the unobserved product characteristics, $\xi_{j,t}$, might be correlated with the number of ad exposures, $N^a_{i,j,t}$ (i.e., $E(\xi_{j,t}|N^a_{i,j,t}) \neq E(\xi_{j,t})$). Second, the unobserved individual characteristics, υ_i, might be correlated with $N^a_{i,j,t}$ (i.e., $E(\upsilon_i|N^a_{i,j,t}) \neq E(\upsilon_i)$). Recall that we denote by υ_i all the individual-specific parameters.[45]

We discuss each of these problems and their solutions below.

Unobserved product characteristics, $\xi_{j,t}$. The number of ad exposures, $N^a_{i,j,t}$, might be correlated with $\xi_{j,t}$, if, for example, firms tend to send more ads for products with high unobserved product characteristics.[46] Without accounting for the potential correlation between N^a and ξ, the estimates of the advertising parameters would be inconsistent.

The availability of individual-level data can resolve this type of endogeneity problem if, in such data sets, N^a varies not only across products but also across individuals. Then, even if the variation in products' market shares is swept away by estimates of the unobserved product characteristics (fixed effects), $\xi_{j,t}$, the advertising parameters can still be estimated. We follow this approach and estimate the $\xi_{j,t}$ of each one of the 64 shows in our data. This also implies that instead of assuming a functional form for the joint distribution of N^a and ξ, we estimate the correlation between them nonparametrically (see discussion in Arellano, 2003).[47]

These estimates are consistent as long as the advertising parameters (which are estimated by the variation across individuals) are consistent.

Unobserved individual characteristics, υ_i. Firms tend to send their ads to consumers who, *a priori*, have a higher tendency to consume the promoted product. In Section 6, we show that such a targeting strategy is both optimal and observed in our data. This strategy leads to a correlation between $N^a_{i,j,t}$ and υ_i. Without accounting for the potential correlation between N^a and υ, the estimates of the advertising parameters would be inconsistent.

[44] We do not test the myopic assumption because doing so has a large computational cost.

[45] For example, υ_i includes υ^x_i, $\upsilon_{i,j}$, and $\upsilon^m_{i,j}$.

[46] A rationale for such a spending strategy was introduced by Milgrom and Roberts (1986). Their theory implies that in equilibrium there is a positive correlation between ad intensity and product quality.

[47] The type of endogeneity problem discussed here is somewhat similar to the one that arises from the correlation between unobserved product characteristics and prices in the differentiated product literature (see, e.g., Berry, Levinsohn, and Pakes, 1995). However, unlike exposure to advertising, prices vary across products and markets but not across individuals in each market. Thus, even when microdata are available, in order to estimate price elasticities, one needs to make some assumptions about the joint distribution of the unobserved product characteristics and prices and to use instruments.

The standard way to resolve this problem is to account for this correlation by including in the estimation the joint distribution of N^a and υ (Arellano, 2003). We follow this approach, which is relatively easy to implement in our application. Specifically, in our application,

$$N^a_{i,j,t} = \sum_{\tau=1}^{t-1} a_{j,t,\tau} I\{y_{i,\tau} = j\}, \tag{16}$$

where the binary variable $a_{j,t,\tau}$ equals one if an ad for the show on network j at time t was aired in time slot τ. Notice that $a_{j,t,\tau}$ is not individual specific and that the variations in $N^a_{i,j,t}$ across individuals result from the heterogeneity in their viewing choices. The critical aspect is that the viewing choices depend on υ_i. "Targeting" means that firms choose $a_{j,t,\tau}$ so that their ads would reach consumers whose tastes, υ_i, fit the promoted show's attributes well. This leads to the endogeneity problem.

Equation (16) implies that $N^a_{i,j,t}$ is a function of $\{y_{i,t-1}, \ldots, y_{i,1}\}$ and $a_{j,t,\tau}$ (i.e., $N^a_{i,j,t}(\{y_{i,t-1}, \ldots, y_{i,1}\}, a_{j,t})$, where $a_{j,t}$ is a vector that comprises $a_{j,t,\tau=1}, \ldots, a_{j,t,\tau=t-1}$). Thus, the joint probability of $N^a_{i,j,t}$ and υ_i depends (through $a_{j,t}$) on the joint probability of $\{y_{i,t-1}, \ldots, y_{i,1}\}$ and υ_i. Let f_υ denote the density function of υ, f^h_{t-1} the history probability up to period $t-1$ conditioned on υ (i.e., $f^h_{t-1}(\{y_{i,t-1}, \ldots, y_{i,1}\}|\upsilon)$), and f_1 the choice probability at period t conditioned on $N^a_{i,j,t}$ and υ. The joint probability of $\{y_{i,t-1}, \ldots, y_{i,1}\}$ and υ is equal to

$$f^h_{t-1}(\{y_{i,t-1}, \ldots, y_{i,1}\}|\upsilon) f_\upsilon(\upsilon). \tag{17}$$

It implies that individuals with different choice histories (and thus different $N^a_{i,j,t}$) are likely to have different υs. In order to resolve the endogeneity problem, one needs to account for these differences in υ across individuals with different $N^a_{i,j,t}$. Thus, in the estimation, we multiply the choice probability in period t with the joint probability in (17):

$$f_1(y_{i,t}|N^a_{i,j,t}(\{y_{i,t-1}, \ldots, y_{i,1}\}, a_{j,t,\tau}), \upsilon) f^h_{t-1}(\{y_{i,t-1}, \ldots, y_{i,1}\}|\upsilon) f_\upsilon(\upsilon). \tag{18}$$

Equation (18) decomposes the correlation between N^a and υ into its two sources: a behavioral effect on $y_{i,t}$ as captured in the choice probability f_1, and a spurious effect due to strategic targeting of advertising as captured in the joint probability $f^h_{t-1} f_\upsilon$. Specifically, if $a_{j,t}$ was chosen strategically (in order to locate individuals whose υ fits the attributes of the promoted show), then $N^a_{i,j,t}$ would be high for individuals whose υ fits the attributes of the promoted show well. Thus, for example, if the correlation between $N^a_{i,j,t}$ and $y_{i,t}$ were fully explained by the correlation between $N^a_{i,j,t}$ and υ, (18) would indicate that $N^a_{i,j,t}$ does not have a behavioral effect on $y_{i,t}$.

Now, equation (18) can be rewritten as

$$f^h_t(\{y_{i,t}, \ldots, y_{i,1}\}|a_{j,t}, \upsilon) f_\upsilon(\upsilon), \tag{19}$$

which is the conditional history probability at period t multiplied by the density function of υ. This implies that in order to resolve the endogeneity problem, one needs to construct the joint probability of the history of choices and υ. This is exactly how we write the likelihood function for the estimation (which is presented in the next subsection).[48]

This discussion illustrates another central advantage of our data set—its ability to effectively tackle the problem of endogeneity of advertising exposures. The reason is that because exposure to advertising depends on television-viewing choices in previous periods, the joint distribution between $N^a_{i,j,t}$ and υ_i is directly related to the model and the data. In contrast, in other typical

[48] (i) Monte Carlo experiments (of a simple version of our model) verify that this approach indeed yields consistent estimates of the advertising parameters. A detailed description of the experiments and their results is available at www.tau.ac.il/~rroonn/Papers/Matchmaker.html.

(ii) There is another way to view the endogeneity problem and its solution in our case. Because the exposure to advertising is determined by previous choices, the ad exposure variable is a form of state dependence (see equation (16)). As suggested by Heckman (1981a), the proper solution for such a problem is to integrate out the unobservables, as is done in our likelihood function.

applications, one needs to (i) assume this joint distribution, and (ii) collect additional data on variables that determine exposures to advertising.

Exogenous variation in the determination of $N^a_{i,j,t}$ can improve the precision of the joint distribution of $N^a_{i,j,t}$ and υ_i and thus the precision of the estimates of the advertising parameters.[49] In our setup, exogenous variation comes from the variation in the product attributes across periods and from the different number of networks that compete in each period.[50]

□ **The likelihood function.** We assume that the $\varepsilon_{i,j,t}$ are drawn from independent and identical Weibull (i.e., independent type I extreme value) distributions. As McFadden (1973) illustrates, under these conditions, the conditional choice probability is multinomial logit:

$$f_1(y_{i,t}|y_{i,t-1}, W_{i,t}; \upsilon_i, \omega_{i,t}, \theta) = \frac{\sum_{j=0}^{J}[I\{y_{i,t} = j\} \exp(\overline{u}_{i,j,t}(y_{i,t-1}, W_{i,j,t}, \upsilon_i, \omega_{i,j,t}, \theta))]}{\sum_{j=0}^{J} \exp(\overline{u}_{i,j,t}(y_{i,t-1}, W_{i,j,t}, \upsilon_i, \omega_{i,j,t}, \theta))}, \quad (20)$$

where $W_{i,j,t}$ is a vector of all the variables in the model (that is, product attributes, $x_{j,t}$, observed individuals' characteristics, $z_{i,t}$, and exposure to ads, $N^a_{i,j,t}$), $W_{i,t}$ is the J-element vector whose jth component is $W_{i,j,t}$, $\omega_{i,t}$ is the J-element vector whose jth component is $\omega_{i,j,t}$, θ is the vector of the parameters that are common to all the individuals (including $\xi_{j,t}$),[51] and $\overline{u}_{i,j,t} = u_{i,j,t} - \varepsilon_{i,j,t}$.

Initial conditions. There are five nights in the studied week. For each night, 8:00 is a natural starting point for the dynamic choice process because the national networks do not air any programs between 7:00 and 8:00 p.m. This means, for example, that the Boston affiliate station that airs ABC programming after 8:00 p.m. might broadcast at 7:45 a show that appears at the same time on the NBC affiliate in New York. This feature of the data suggests that the "initial conditions" might not pose a problem in our case. However, we still proceed by solving the initial-conditions problem in the standard way.

For each individual, we observe 12 choices between 8:00 and 11:00 p.m. for each of the five nights of the week. The 8:00 choice probability depends on the 7:45 choice, which is an endogenous variable. It depends on some of the same parameters driving the choices in later periods. Using the 7:45 choice as if it were exogenous would lead to a biased and inconsistent estimator, as described in Heckman (1981b). A solution to this initial-conditions problem is to endogenize the 7:45 choice.

Implementing this solution to the initial-conditions problem is not trivial, because our data do not specify which channel is watched when viewing occurs at 7:45. Nielsen does not record the 7:45 network choices because between 7:00 and 8:00 p.m., the affiliate stations broadcast local programming. We know only if the TV was off ($y_{i,t-1} = 0$) or on ($y_{i,t-1} \neq 0$). When $y_{i,t-1} = 0$, the lagged choice is observed, and we account for the dependence of $y_{i,t-1}$ on the unobservables υ_i as follows:

[49] See Carrasco (2001).

[50] For example, the ad for the ABC show *20/20* (Friday at 10:00 p.m.) appears during the show *Coach* (Tuesday at 9:30 p.m.). The show *20/20* is a news magazine that competes with a sports program on CBS and an action drama on NBC; furthermore, Fox does not air national programming during this time slot. *Coach* is a sitcom that competes with a romantic movie on CBS, a sitcom on NBC, and an action movie on Fox. Notice that because Fox does not air national programming after 10 p.m., there are periods with four competing networks and periods with three competing networks.

Berry (1994) and Berry, Levinsohn, and Pakes (1995) were the first to employ product attributes of competing firms as exogenous sources of variation that can assist in resolving endogeneity problems in discrete-choice models, although the justification for using such variables is quite different there.

[51] That is, $\theta = \{\xi_{j,t}, \beta_{Gender}, \beta_{Age0}, \beta_{Age1}, \beta_{Age2}, \beta_{Family}, \beta_{RaceIncome}, \beta_{Sitcom}, \beta_{AD}, \beta_{RD}, \beta_{NewsMagazine}, \beta_{Sports}, \delta^z, \delta^x, \delta_{First15}, \delta_{Last15}, \delta_{Cont}, \delta_{InProgress}, \delta_{Hour}, \delta_0, \gamma, \rho_1, \rho_2, \varsigma^m_{New}, \varsigma^m_{Veteran}, \varsigma^a\}$.

$$f_1(y_{i,t}|y_{i,t-1}, W_{i,t}; \upsilon_i, \omega_{i,t}, \theta)\frac{\exp(\overline{u}_{i,0,t-1}(z_{i,t-1}; \upsilon_i, \theta_0))}{\sum_j \exp(\overline{u}_{i,j,t-1}(z_{i,t-1}; \upsilon_i, \theta_0))}, \tag{21}$$

where the second term is the logit probability for 7:45. Because viewers are exposed to local programming, we do not observe the shows' attributes for this time slot. Thus, for $j > 0$, the utility at 7:45 is only $(\xi_{j,t} + \upsilon_{i,j} + \varepsilon_{i,j,t})$.[52]

When $y_{i,t-1} \neq 0$, we not only account for the dependence of each possible unobserved viewing choice $\widehat{y}_{i,t-1}$ on υ_i but also integrate over all these unobserved choices:

$$\sum_{\widehat{y}_{i,t-1}=1}^{4} f_1(y_{i,t}|\widehat{y}_{i,t-1}, W_{i,t}; \upsilon_i, \omega_{i,t}, \theta)\frac{\exp(\overline{u}_{i,\widehat{y}_{i,t-1},t-1}(\upsilon_i, \theta_1))}{\sum_j \exp(\overline{u}_{i,j,t-1}(\upsilon_i, \theta_1))}. \tag{22}$$

Goettler and Shachar (2001), who introduce this solution to the 7:45 unobserved choices, also discuss the identification of the relevant parameters. Specifically, because we estimate a single $\xi_{j,t}$ for all the time slots of a show, and because each show spans at least two such time slots, one can identify the $\xi_{j,7:45}$ parameters.[53]

To summarize: we solve the initial-conditions problem by structurally accounting for the dependence of the 7:45 choices on all relevant unobservables, including those that appear in other time periods, and integrating them (below) from the history probability.

History probability. After including the 7:45 time slots in the analysis, we have 13 choices for each night and $T = 65$. Let $f_{8:00}(y_{i,t}|y_{i,t-1}, W_{i,t}; \upsilon_i, \omega_{i,t}, \theta)$ represent the 8:00 choice probabilities above, and $y_i = \{y_{i,1}, \ldots, y_{i,65}\}$ denote individual i's history of choices for the entire week. Recall that $\varepsilon_{i,j,t}$ are independent over time. Thus, the history probability is then

$$f_2(y_i|W_i; \upsilon_i, \omega_i, \theta) = \prod_{d=1}^{5}\left[f_{8:00}(y_{i,13d-11}|y_{i,13d-12}, W_{i,13d-11}; \upsilon_i, \omega_{i,13d-11}, \theta)\prod_{t=(13d-10)}^{13d} f_1(y_{i,t}|y_{i,t-1}, W_{i,t}; \upsilon_i, \omega_{i,t}, \theta)\right], \tag{23}$$

where W_i is the T-element vector whose tth component is $W_{i,t}$ and ω_i is defined accordingly.

Integrating out the unobserved ω of the first show on ABC, we get

$$\int_{\widetilde{\omega}_1} f_2(y_i|W_i; \upsilon_i, (\widetilde{\omega}_1, \ldots, \widetilde{\omega}_{64}), \theta)\phi(\widetilde{\omega}_1)\, d\widetilde{\omega}_1,$$

where $\phi(\widetilde{\omega}_1)$ is the standard normal density function. Repeating this integration for the other 63 shows in the week gives us the history probability unconditional on ω_i, $f_3(y_i|W_i; \upsilon_i, \theta)$. Specifically,

$$f_3(y_i|W_i; \upsilon_i, \theta) = \int_{\widetilde{\omega}_{64}} \cdots \int_{\widetilde{\omega}_1} f_2(y_i|W_i; \upsilon_i, (\widetilde{\omega}_1, \ldots, \widetilde{\omega}_{64}), \theta)\phi(\widetilde{\omega}_1)d\widetilde{\omega}_1 \cdots \phi(\widetilde{\omega}_{64})\, d\widetilde{\omega}_{64}.$$

Recall that for any individual, $\omega_{i,j,t}$ is constant across all time slots of a specific show, and it is independent across shows. In practice, because $\omega_{i,j,t}$ is show specific, none of the integrals should include the entire history.[54]

[52] This means that we estimate a specific $\xi_{j,t}$ for each network-night combination for 7:45. The parameter vector θ_1 (presented immediately) consists of these 20 parameters. The vector θ_0 includes all of the parameters in θ_1 and the 7:45 specific γ parameters.

[53] For example, given $\xi_{j,8:15}$, and the fact that $\xi_{j,8:00}$ equals $\xi_{j,8:15}$, shows with a larger than expected audience at 8:00 probably had a larger lead-in audience from 7:45. This large lead-in audience translates into a higher $\xi_{j,7:45}$.

[54] Each integral includes only the time slots during which the relevant show is aired. For example, on Wednesday between 10:00 and 11:00 p.m., each of the three major networks airs a one hour show. Thus, for these time slots, the

Notice that f_3 is the history probability conditional on υ_i discussed in the last subsection (where it was termed f^h). In order to resolve the endogeneity problem (as in equation (18)), we need to multiply it by the density function of υ.

The density function of υ_i, denoted by f_υ, is assumed to be discrete. Specifically, $\upsilon_i = \upsilon_k$ with probability $\frac{\exp(\pi_k)}{\sum_k \exp(\pi_k)}$ for all $k = 1, \ldots, K$. This means that we allow the population to be divided into K different unobserved segments. The number of types K is determined based on various information criteria. The discrete distribution has at least two advantages over the continuous specification. First, it does not assume any functional form. Second, it easily allows correlation between the unobserved parameters—for example, it might be the case that a person who likes NBC ($\upsilon_{i,NBC}$ is high) also knows a lot about this network's shows ($\upsilon^m_{i,NBC}$ is high).

Integrating out the unobserved individual-specific parameters, υ_i, we get the marginal probability

$$f_4(y_i|W_i;\theta') = \sum_{k=1}^{K} f_3(y_i|W_i;\upsilon_k,\theta)\frac{\exp(\pi_k)}{\sum_{k=1}^{K}\exp(\pi_k)}. \tag{24}$$

where θ' includes θ, the υ_ks, and the πs.

The likelihood function is

$$L(\theta') = \prod_{i=1}^{I} f_4(y_i|W_i;\theta'). \tag{25}$$

□ **Simulating the marginal probability.** Because $\omega_{i,j,t}$ is normally distributed, the integrals of $f_3(y_i|W_i;\ \upsilon_i,\ \theta)$ do not have a closed-form solution. Consistent and differentiable simulation estimators of $f_3(\cdot)$ and $f_4(\cdot)$ are

$$\hat{f}_3(y_i|W_i;\upsilon_i,\theta) = \frac{1}{R}\sum_{r=1}^{R} f_2(y_i|W_i;\upsilon_i,\omega_r,\theta) \tag{26}$$

$$\text{and } \hat{f}_4(y_i|W_i;\theta') = \sum_{k=1}^{K} \hat{f}_3(y_i|W_i;\upsilon_k,\theta)\frac{\exp(\pi_k)}{\sum_{k=1}^{K}\exp(\pi_k)}, \tag{27}$$

where the ωs are randomly drawn from the standard normal distribution. The maximum simulated likelihood (MSL) estimator is then

$$\widehat{\theta'}_{MSL} = argmax \sum_{i=1}^{I} \log\left(\hat{f}_4(y_i|W_i;\theta')\right). \tag{28}$$

As explained in McFadden (1989) and Pakes and Pollard (1989), the R variates for each individual's ωs must be independent and remain constant throughout the estimation procedure. A drawback of using MSL is the bias of θ'_{MSL} due to the logarithmic transformation of $f_3(\cdot)$. Despite this bias, the estimator obtained by MSL is consistent if $R \to \infty$ as $I \to \infty$, as detailed in Proposition 3 of Hajivassiliou and Ruud (1994). To attain negligible inconsistency, Hajivassiliou (1997) suggests increasing R until the expectation of the score function is zero at θ'_{MSL}.[55] In our case, this is achieved at $R = 400$.

integration is only over three unobserved ωs. Indeed, the largest number of integrals for each time slot is 13. We use this feature to rewrite the history probability in order to minimize the number of integrals for each time slot.

[55] We simulate all stochastic components of the model to construct an empirical distribution of the score function at $\widehat{\theta'}_{MSL}$. A quadratic form of this score function is asymptotically distributed χ^2 with degrees of freedom equal to the number of parameters estimated.

In order to reduce the variance of $\hat{f}_3(\cdot)$, we employ *importance sampling* as described in the Monte Carlo literature (see Rubinstein, 1981). Our importance sampler is similar to the one used in BLP (1995). We draw the ωs from a multivariate normal approximation of each person's posterior distribution of ω, given some preliminary MSL estimate of θ', and appropriately weight the conditional probabilities to account for the oversampling from regions of z that lead to higher probabilities of i's actual choices. For $R = 400$, we find that importance sampling reduces the RMSE of $\hat{f}_4(\cdot)$ to about 0.42 the size of the root mean square error (RMSE) when not using importance sampling.[56]

□ **Identification.** We start by considering the identification of a model under the assumption that the individual is fully informed (that is, under the assumption that $\frac{1}{\varsigma^m_{i,j,t}} = 0$ for all i and j). This discussion illustrates which parameters can be identified without the structural restrictions imposed by the full model and the additional variables introduced by this model.

Utility parameters. The β parameters are identified by the correlation between $x_{j,t}z_i$ and viewer choices. The unobserved tastes for show categories, the υ^x_i parameters, are identified by the conditional viewer choice histories over show types. The unobserved product attribute $\xi_{j,t}$ is identified by the conditional aggregate show ratings. The parameters of the h function are identified by the conditional state dependence—that is, by the share of viewers who remain with an alternative over two sequential time slots, conditioning (among other elements) on $x_{j,t}$. The parameter $\upsilon_{i,j}$ is identified by the conditional viewer choice histories over networks. Notice that a positive $\upsilon_{i,j}$ leads individual i to view shows on network j even when those shows do not fit her preferences well.[57] The conditional correlation between the number of advertising exposures and viewing choices identifies the parameters of the direct effect of ads on the utility, denoted by ρ.

Information set parameters. The partial information model imposes some restrictions on the parameters and introduces new explanatory variables. These identify the information set parameters. We start by discussing the estimation of the prior distribution parameters and then present the identification of the signals' parameters.

Prior distribution. Recall that the estimation of the prior distribution parameters was already discussed in Section 3. Once the parameters β and ξ are identified, we also have an estimate of all the variables that are a function of them, namely $u^{att}_{i,j,t}$, $\mu_{i,j}$, and $\varsigma^{\mu}_{i,j}$.

It is worth clarifying the identifying source of this distinction between $\upsilon_{i,j}$ and $\mu_{i,j}$. In the model, we set $\hat{\mu}_{i,j} = \frac{1}{T}\sum_{t=1}^{T}\widehat{u}^{att}_{i,j,t} = \frac{1}{T}\sum_{t=1}^{T}\widehat{\xi}_{j,t} + (\frac{1}{T}\sum_{t=1}^{T}x_{j,t})\hat{\beta}_i$. Thus, the identification of $\mu_{i,j}$ is based on an explanatory variable that does not exist in a model where consumers are fully informed. Specifically, this variable is the mean offering of each network (for example, $\frac{1}{T}\sum_t x_{j,t}$ for network j).

[56] The RMSE of $\hat{f}_4(y_i|W_i;\theta')$ is computed using N_R sets of R draws as

$$RMSE(R) = \left[\frac{1}{N_R}\sum_{n=1}^{N_R}\frac{(\hat{f}_{4,n}(y_i|W_i;\theta) - f_{4,true})^2}{f_{4,true}}\right]^{0.5}, \tag{29}$$

where $f_{4,true}$ represents the true value. Because this true value is not computable, we evaluate $\hat{f}_4(\cdot)$ using $R = 2^{20}$ Monte Carlo draws and take this to be the true value. Any reduction in the variance of the estimator for $\hat{f}_4(\cdot)$ reduces the bias and variance of the estimator of θ'. Quantifying the magnitude of this reduction is of interest. To our knowledge, constructing the empirical distribution of $\widehat{\theta}'_{MSL}$ via a bootstrapping method is the only way to proceed. Unfortunately, the cpu time required to compute $\widehat{\theta}'_{MSL}$ prohibits us from pursuing this goal.

[57] As discussed in the literature, there are various sources of identifying the state-dependence parameters separate from υ (Heckman, 1981a; Chamberlain, 1993; Shachar, 1994). The outside alternative provides us with an additional identifying source. When turning on the television, the individual's "state" (lagged choices) does not attach her to any network. Thus, her viewing choice is influenced by υ (and show characteristics), but not by the state-dependence parameters.

TABLE 3A The Sizes of the Segments

Parameter	$\pi_{k=1}$	$\pi_{k=2}$	$\pi_{k=3}$	$\pi_{k=4}$	$\pi_{k=5}$	$\pi_{k=6}$
Mean	0	.804	−.243	.927	.633	.435
	(-)	(.225)	(.238)	(.211)	(.219)	(.215)
Size of segment	.100	.224	.079	.253	.189	.155

Product-specific signals. The parameters of the product-specific signals are ς^m and ς^a. The sum of the precision of all the product-specific signals (advertising and miscellaneous), $\varsigma^a N^a_{i,j,t} + \varsigma^m_{i,j,t}$, enters $\overline{u}_{i,j,t}$ (and, thus, the likelihood) only through $\lambda_{i,j,t}$ and $\sigma^{\omega}_{i,j,t}$. Furthermore, neither ς^a nor ς^m enters the likelihood in any other form.

The dependence of $\lambda_{i,j,t}$ and $\sigma^{\omega}_{i,j,t}$ on $\varsigma^a N^a_{i,j,t} + \varsigma^m_{i,j,t}$ leads to various identifying factors. For example, as discussed in Section 3, if $\frac{1}{\varsigma^a N^a_{i,j,t}+\varsigma^m_{i,j,t}} = 0$, then $\lambda_{i,j,t} = 1$, and the choice of a product is not a function of the multiproduct firm profile $\mu_{i,j}$ (see equation (12)). Thus, any evidence that choices depend on the firm profile $\mu_{i,j}$ implies that the estimate of $\frac{1}{\varsigma^a N^a_{i,j,t}+\varsigma^m_{i,j,t}}$ is positive. Furthermore, the larger the correlation between choices and the firm profile $\mu_{i,j}$, the larger the estimate of $\frac{1}{\varsigma^a N^a_{i,j,t}+\varsigma^m_{i,j,t}}$ (i.e., the smaller the $\varsigma^a N^a_{i,j,t} + \varsigma^m_{i,j,t}$). Notice that this source of identification relies on *observed* product and firm attributes.

Now, given that $\varsigma^a N^a_{i,j,t} + \varsigma^m_{i,j,t}$ is identified, variation in $N^a_{i,j,t}$ enables us to empirically distinguish between ς^a and $\varsigma^m_{i,j,t}$.[58] Notice that these identifying factors are related to each of the implications of informative advertising that was discussed in Section 3 (e.g., consumption deterrence and matching).

As seen, each of these effects of advertising through the information set is different from the direct effect of advertising on the utility, allowing us to empirically distinguish between informative and persuasive advertising. Furthermore, as shown earlier in Section 3, the differences in the effects of advertising exposure through these two avenues are not only econometric but behavioral. In other words, identification is driven by the data rather than by model structure alone. One way to see this intuitively is to notice that if $\varsigma^a > 0$, then the effect of $X_{j,t}$ on choices is stronger when N^a is higher, that is, it resembles an interaction term between $X_{j,t}$ and $N^a_{i,j,t}$.[59] In other words, consider the following reduced-form regression: $U_{i,j,t} = \cdots + (1 + \gamma N^a)\beta_i X_{j,t} + \cdots$, where $\gamma > 0$ would reveal an informative effect of advertising. The magnitude of this "interaction term" identifies the precision of the advertising signals. In contrast, the persuasive effect of advertising does not imply any such interaction between N^a and show characteristics.

5. Results

■ The behavioral implications of informative advertising can be tested directly. Indeed, in Section 2, we presented nonstructural evidence on the matching role of advertising and, later, in Section 6, we use the estimates of the model to directly examine the consumption-deterring effect. At the same time, the structural estimation, by embodying all the elements and the restrictions of the model, offers a precise measure of the informative effect of advertising. This section presents the results of the structural estimation. The key parameter of interest is ς^a. If the estimate of ς^a is positive, advertising affects choices through the information set.

Tables 3A–3I present the estimates of the utility and the information set parameters (including the parameter of interest, ς^a, which is discussed last).[60]

[58] For example, the estimate of ς^a would be positive if one finds in the data that an increase in $N^a_{i,j,t}$ leads to an increase in $\lambda_{i,j,t}$ (i.e., a reduction in the effect of $\mu_{i,j}$ on choices).

[59] For simplicity, ignore the effect on the variances.

[60] The integrals in $f_3(y_i|W_i;\, \upsilon_i, \theta)$ are evaluated numerically using importance sampling with 400 points from a pseudorandom sequence as detailed in Section 4. The (asymptotic) standard errors are derived from the inverse of the

TABLE 3B Preferences for Show Attributes (β and ν)

Cast Demographics					
β_{Gender}	β_{Age0}	β_{Age1}	β_{Age2}	β_{Family}	β_{Race}
.258	1.183	.847	0	.508	−1.004
(.055)	(.111)	(.093)	(-)	(.127)	(.317)

Show Genre	Sitcom	Action Drama	Romantic Drama	Sports
Teens	0	0	0	0
	(-)	(-)	(-)	(-)
Generation-X	−1.040	−.761	−.003	0.159
	(.234)	(.264)	(.301)	(.339)
Baby Boomer	−1.128	−.597	−.011	.100
	(.226)	(.250)	(.284)	(.335)
Older	−1.578	−.609	−.154	.179
	(.247)	(.271)	(.309)	(.352)
Female	0.276	.373	.738	−.409
	(.102)	(.121)	(.150)	(.150)
Income	−.045	−.588	−1.378	.257
	(.285)	(.333)	(.340)	(.408)
Education	−.257	−.046	−.893	−.167
	(.263)	(.303)	(.362)	(.357)
Family	.218	−.007	.170	−.043
	(.143)	(.141)	(.175)	(.182)
First segment	0	0	0	0
	(-)	(-)	(-)	(-)
Second segment	−1.863	−2.218	−2.138	1.112
	(.286)	(.254)	(.496)	(.558)
Third segment	.456	−2.483	−2.997	−1.259
	(.243)	(.265)	(.585)	(.578)
Fourth segment	−.206	−1.173	.423	.301
	(.236)	(.242)	(.378)	(.553)
Fifth segment	.840	−.337	−1.266	.823
	(.231)	(.245)	(.425)	(.565)
Sixth segment	−.619	−1.340	−.939	−.578
	(.216)	(.210)	(.378)	(.558)

TABLE 3C The Mean and Standard Deviation of the Shows' Fixed Effects

	ABC	CBS	NBC	Fox	All networks
Mean	−4.412	−4.866	−4.095	−5.200	−4.593
Standard deviation	2.341	3.038	2.066	3.570	2.747

The results are for a model with six segments ($K = 6$). The number of unobserved segments was determined by minimizing the Bayes information criterion. The largest segment consists of about 25% of the population, whereas the proportion of the smallest segment is about 8%. The sizes of the other segments are 0.22, 0.19, 0.15, and 0.10; the π parameters, which determine the sizes of the segments, are reported in Table 3A.

simulated information matrix. The reported standard errors, therefore, neglect any additional variance due to simulation error in the numerical integration.

TABLE 3D Individual-Brand Match Parameters ($\nu_{i,j}$)

		ABC	CBS	NBC	Fox
Segment Number	1	0	0	0	0
	2	.339	.271	.258	.560
		(.233)	(.196)	(.208)	(.212)
	3	1.199	1.365	.667	.278
		(.245)	(.206)	(.216)	(.381)
	4	.659	−.285	.116	.679
		(.203)	(.199)	(.184)	(.181)
	5	.151	.406	.055	−.547
		(.222)	(.192)	(.184)	(.256)
	6	.488	.296	.171	.555
		(.209)	(.186)	(.185)	(.195)

TABLE 3E Preference for Outside Alternatives (γ and $\nu_{i,Out,t}$)

Parameter	Estimate (Standard Error)	Parameter	Estimate (Standard Error)	Parameter	Estimate (Standard Error)
γ_{Basic}	0.266	$\nu_{k=1,Out,9-10PM}$	0.000	$\gamma_{8:00}$	0.000
	(.049)		(—)		(—)
$\gamma_{Premium}$	0.350	$\nu_{k=1,Out,10-11PM}$	0.000	$\gamma_{8:15}$	−0.466
	(.051)		(—)		(.116)
γ_{All}	−0.723	$\nu_{k=2,Out,9-10PM}$	0.110	$\gamma_{8:30}$	−0.406
	(.115)		(.270)		(.120)
γ_{Same}	−0.643	$\nu_{k=2,Out,10-11PM}$	1.737	$\gamma_{8:45}$	−0.497
	(.073)		(.270)		(.150)
γ_{Teens}	0.000	$\nu_{k=3,Out,9-10PM}$	0.314	$\gamma_{9:00}$	−0.574
	(—)		(.245)		(.230)
$\gamma_{Generation\text{-}X}$	−0.953	$\nu_{k=3,Out,10-11PM}$	1.784	$\gamma_{9:15}$	−0.565
	(.200)		(.285)		(.243)
$\gamma_{BabyBoomer}$	−1.060	$\nu_{k=4,Out,9-10PM}$	0.743	$\gamma_{9:30}$	−0.630
	(.197)		(.223)		(.240)
γ_{Older}	−1.449	$\nu_{k=4,Out,10-11PM}$	2.270	$\gamma_{9:45}$	−0.489
	(.212)		(.256)		(.249)
γ_{Female}	0.027	$\nu_{k=5,Out,9-10PM}$	0.825	$\gamma_{10:00}$	−1.268
	(.087)		(.225)		(.301)
γ_{Income}	−0.494	$\nu_{k=5,Out,10-11PM}$	2.184	$\gamma_{10:15}$	−1.210
	(.234)		(.249)		(.317)
$\gamma_{Education}$	−0.487	$\nu_{k=6,Out,9-10PM}$	0.180	$\gamma_{10:30}$	−0.922
	(.226)		(.217)		(.316)
γ_{Family}	0.051	$\nu_{k=6,Out,10-11PM}$	1.543	$\gamma_{10:45}$	−1.079
	(.112)		(.239)		(.323)

□ **Utility parameters.** We briefly summarize the estimates of the utility parameters here. These are consistent with the results of previous studies.

The consumer-product match parameters are presented in Table 3B. Like previous studies, we find that viewers prefer shows whose cast demographics (age, gender, family, race) are similar to their own, and that preference heterogeneity over show genres depends on viewers' observed and unobserved characteristics. For example, female viewers like romantic dramas the most and sports events the least.

We have estimated the unobserved product attribute for each of the 64 shows in our data (except one, for normalization). For brevity, Table 3C presents only the mean and standard deviation of the fixed-effect parameters $\xi_{j,t}$ for each of the networks. On average, the mean is −4.593 and the standard deviation is 2.747. The show with the highest $\xi_{j,t}$ is the NBC drama *ER*.

TABLE 3F State Dependence Parameters

Parameter	Estimate
δ_{Sitcom}	0.619
	(.131)
$\delta_{ActionDrama}$	0.870
	(.126)
$\delta_{RomanticDrama}$	0.545
	(.128)
$\delta_{NewsMagazine}$	0.000
	(—)
δ_{Sport}	−0.655
	(.151)
$\delta_{k=1}$	2.622
	(.137)
$\delta_{k=2}$	2.243
	(.134)
$\delta_{k=3}$	1.944
	(.126)
$\delta_{k=4}$	2.517
	(.119)
$\delta_{k=5}$	2.794
	(.126)
$\delta_{k=6}$	1.329
	(.118)
δ_{Basic}	−0.428
	(.049)
$\delta_{Premium}$	−0.384
	(.053)
δ_{Female}	0.079
	(.041)
δ_{Family}	0.018
	(.053)
δ_{Teens}	0.000
	(—)
$\delta_{Generation\text{-}X}$	0.036
	(.085)
$\delta_{BabyBoomer}$	0.017
	(.081)
δ_{Older}	−0.012
	(.089)
$\delta_{Continuation}$	0.857
	(.137)
δ_{Out}	0.693
	(.087)
$\delta_{First15}$	−0.213
	(.099)
δ_{Last15}	0.491
	(.145)
δ_{Hour}	−0.411
	(.090)
$\delta_{In\ Progress}$	−0.331
	(.079)

Table 3D shows that there is significant unobserved heterogeneity in the individual-network match parameters $\upsilon_{i,j}$.

Table 3E presents the parameters of the outside utility. It demonstrates the significant observed and unobserved heterogeneity with respect to the outside utility. For example, the outside alternative is more attractive to younger viewers than to older viewers.

TABLE 3G Parameters of the Direct Effect of Ads on the Utility

		$\rho_{1,MT}$	$\rho_{1,WF}$	$\rho_{2,MT}$	$\rho_{2,WF}$
Segment Number	1	0.732	0.430	0.056	0.030
		(0.368)	(0.264)	(0.119)	(0.078)
	2	2.797	1.151	−0.515	−0.062
		(0.483)	(0.289)	(0.153)	(0.089)
	3	0.904	0.236	−0.128	0.065
		(0.391)	(0.279)	(0.134)	(0.090)
	4	0.830	0.827	−0.115	−0.058
		(0.279)	(0.244)	(0.087)	(0.077)
	5	0.692	0.729	−0.052	−0.086
		(0.322)	(0.265)	(0.105)	(0.085)
	6	0.606	0.592	0.066	0.001
		(0.351)	(0.265)	(0.127)	(0.090)

	ρ_1	ρ_2		ρ_1	ρ_2
Teens	0.000	0.000	Female	0.024	0.004
	(—)	(—)		(.075)	(.024)
GenX	−0.126	−0.007	Income	0.002	−0.050
	(.173)	(.057)		(.185)	(.053)
Boomer	−0.145	0.010	Education	−0.276	0.021
	(.171)	(.056)		(.204)	(.064)
Older	0.046	−0.019	Family	−0.030	−0.005
	(.183)	(.060)		(.105)	(.034)

TABLE 3H Precision of the Prior (ς^{μ})

		ABC	CBS	NBC	Fox
Segment Number	1	0.226	0.116	0.162	0.075
	2	0.344	0.093	0.136	0.076
	3	0.139	0.150	0.107	0.062
	4	0.245	0.116	0.205	0.095
	5	0.239	0.127	0.128	0.062
	6	0.214	0.119	0.159	0.079
Average		0.251	0.116	0.156	0.078

TABLE 3I Precision of the Miscellaneous Signals (ς^{m})

		ABC	CBS	NBC	Fox
Segment Number	1	.687	0.063	0.493	0.049
		(.346)	(.020)	(.181)	(.019)
	2	1.279	0.059	0.270	0.064
		(.778)	(.016)	(.097)	(.028)
	3	0.345	0.111	0.317	0.071
		(.112)	(.037)	(.109)	(.052)
	4	0.473	0.160	0.163	0.034
		(.169)	(.054)	(.043)	(.012)
	5	1.011	0.070	0.715	0.051
		(.448)	(.021)	(.293)	(.021)
	6	0.299	0.100	0.245	0.060
		(.088)	(.028)	(.057)	(.020)
Average		0.740	0.098	0.349	0.052

Like previous studies, our estimates of the state-dependence parameters (Table 3F) are large. For example, the probability of watching the first time slot of a show conditioned on watching the previous show on that network is 41.8%.[61] State dependence is lower for viewers with access to cable channels, and higher for female viewers compared with males.

As explained in Section 4, we have solved the initial-conditions problem by accounting for the dependence of the unobserved 7:45 choices on the individual-specific unobservables, and then integrating over these unobserved choices.[62]

The direct effect of advertising (ρ). The two sets of ρ parameters (for Monday–Tuesday and Wednesday–Friday) are presented in Table 3G. We focus on the Wednesday through Friday parameters here.

The utility of viewers is a positive function of exposures to advertising. On average, the first exposure increases the probability of watching a show by 39.1%, the second exposure increases the viewing probability by an additional 27.6%, and the effect of the third exposure falls to an increase of 16.2%.[63] These results also illustrate the wear-out effect of advertising (i.e., diminishing returns). Both the effect of advertising and the wear-out effect differ across viewer segments, with the advertising effect being weakest for the smallest viewer segment, 3, and the wear-out effect being absent for segments 1, 3, and 6. Whereas there are significant differences across consumers from unobserved sources, the observed consumers' characteristics do not have a significant effect on the ρ parameters.

As discussed in Section 4, targeting strategies by networks creates an endogeneity concern that we account for in the estimation by specifying the joint distribution of the ad exposure variable and the unobservables. The *estimated* correlation between the two variables in the joint distribution can be illustrated by a simple example. Consider the NBC sitcom *Seinfeld*. The expected unobserved taste for sitcoms is equal to $-.521$ for the 1088 individuals who were not exposed to any ad for this show (i.e., $E(\upsilon_i^{Sitcom}|N^a_{i,Seinfeld} = 0) = -.521$). For the 361 individuals who were exposed to one ad and for the 226 who were exposed to two ads or more, the expected values are $-.161$ and $-.003$, respectively. This means that the targeting strategy of NBC ensures that individuals who like sitcoms are indeed more likely to be exposed to ads for *Seinfeld* than those who like this genre less, and that our estimation accounts for this joint distribution and corrects for this correlation.

Ignoring the endogeneity problem would result in inconsistent estimates of the ρ parameters. To illustrate this bias, we have also estimated a model without unobserved heterogeneity (and thus without accounting for the joint distribution of the unobservables and the ad exposure variable). The effect of ad exposure on choices in this case is 47.5% for the first exposure (compared with 39.1% above), and an additional 30.2% for the second exposure.[64]

Another type of endogeneity might arise from the correlation between ad intensity and unobserved product attributes, $\xi_{j,t}$. As discussed in Section 4, we estimate a $\xi_{j,t}$ for each one of the 64 shows in the week and, thus, directly estimate this correlation. In our data, this correlation is small (-0.024) and insignificantly different from zero.

□ **Information set parameters.** Individuals have three sources of information: (i) the distribution of product attributes within each multiproduct firm; (ii) miscellaneous product-specific signals (word of mouth, media coverage, previous experience); and (iii) advertising signals. The resulting parameters of the information set are $\varsigma^{\mu}_{i,j}$ (the precision of the prior distribution), $\varsigma^{m}_{i,j}$ (the precision of the miscellaneous signal), and ς^{a} (the precision of advertising signals). The estimates are presented in this order below and in Tables 3H and 3I.

[61] The probabilities for each segment are, respectively, 0.364, 0.316, 0.481, 0.425, 0.550, and 0.233.

[62] The parameter estimates for the 7:45 choices are posted at www.tau.ac.il/~rroonn/Papers/Matchmaker.html.

[63] In these calculations, we allow ads to affect choices only through the utility.

[64] In the working paper version of this study (Anand and Shachar, 2001), we compare the results of seven different models with varying degrees of heterogeneity. These results demonstrate that our modelling of consumer preferences is rich enough.

Precision of the prior distribution (ς^{μ}). Unlike $\varsigma^m_{i,j}$ and ς^a, the prior distribution parameters $\varsigma^{\mu}_{i,j}$ are not estimated directly but rather as a function of $\widehat{\xi}_{j,t}$ and $\hat{\beta}_i$.[65] The average $\widehat{\varsigma}^{\mu}_{i,j}$ for each viewer segment and network is presented in Table 3H. The average $\widehat{\varsigma}^{\mu}_{i,j}$ is the highest for ABC and the lowest for Fox ($\widehat{\varsigma}^{\mu}_{i,ABC} = 0.251$, $\widehat{\varsigma}^{\mu}_{i,CBS} = 0.116$, $\widehat{\varsigma}^{\mu}_{i,NBC} = 0.156$, and $\widehat{\varsigma}^{\mu}_{i,Fox} = 0.078$).[66]

Precision of miscellaneous signals (ς^m). The parameter $\upsilon^m_{i,j}$ (which is the unobserved element in ς^m) can be thought of as viewer i's degree of familiarity with the shows on network j. The heterogeneity across individuals in their familiarity level is evident from the estimates ($\hat{\upsilon}^m_{i,j}$ varies from 0.034 to 1.27). Because these signals are product specific, these estimates imply that individuals differ in their prior information about each show even without any exposures to advertising.

On average, viewers are more familiar with shows on ABC and NBC than those of the other networks (the averages are 0.740 for ABC, 0.349 for NBC, 0.098 for CBS, and 0.052 for Fox). These estimates are sensible for the following reasons. The degree of familiarity with a network should be a positive function of (i) the ratings of its shows and (ii) the "age" of its shows (i.e., the number of seasons that the shows were on the air). The reason is that information from both word-of-mouth sources and previous experience tends to be larger for successful and veteran shows.

Thus, one would have expected that viewers would be quite familiar with the shows on ABC, whose shows were both popular and veterans. Although NBC was leading the "ratings race" in 1995, it is not surprising that viewers were a bit less familiar with its shows, because many of its popular shows were new.[67] The low $\hat{\upsilon}^m_{i,j}$ for CBS and Fox are not surprising as well—their average rating lagged that of the other networks, and CBS had additionally introduced many new shows in the fall of 1995.[68]

Precision of advertising signals (ς^a). Unlike previous studies, the model presented here allows the information set to depend on advertising content. Section 2 provides preliminary evidence to support this approach. In the structural estimation, the empirical evidence in favor of this approach rests on whether $\varsigma^a > 0$ or not.

The data support the theory—the estimate of ς^a is 0.2739 with a standard error of 0.0741, and is statistically different from zero at the 1% level. The effect of advertising through the information set is behaviorally important as well. This is illustrated below in several ways.

Because the average $\hat{\varsigma}^m_{i,j}$ across viewers and networks is 0.309, and the average $\hat{\varsigma}^{\mu}_{i,j}$ is 0.150, the precision of a single advertising signal is almost the same as the precision of all other miscellaneous signals and almost twice as large as precision of the prior distribution.

Recall that $\lambda_{i,j,t}$ can be considered as a measure of how well informed an individual is. Table 4 presents this measure as a function of the number of advertising exposures. On average, $\lambda = 0.555$ when the number of exposures is zero. With one exposure, λ increases to 0.788. The breakdown of

[65] Specifically, because our estimate of the utility attribute is $\hat{u}^{att}_{i,j,t} = \hat{\xi}_{j,t} + x_{j,t}\hat{\beta}_i$, it follows that (if we were not treating the two parts of the night, 8:00–10:00 and 10:00–11:00 p.m., separately), $\widehat{\varsigma}^{\mu}_{i,j} = [\frac{1}{T-1}\sum_t(\hat{u}^{att}_{i,j,t} - \frac{1}{T}\sum_t \hat{u}^{att}_{i,j,t})^2]^{-1}$. We calculate $\widehat{\varsigma}^{\mu}_{i,j}$ for each part of the night accordingly.

[66] The finding for Fox may seem surprising, because this network appears to offer the most homogeneous profile of shows: many Generation-X dramas, and no sitcoms or news magazines. However, recall that the attribute utility is a function of both $\xi_{j,t}$ and $x_{j,t}$. Whereas the variance in $x_{Fox,t}$ is indeed the lowest among the four networks, the variance of $\xi_{Fox,t}$ is the highest.

[67] Even though NBC enjoyed the highest average rating (followed by ABC in second place) during the fall season of 1995, it was only third in the ratings race during the 1994 season (behind ABC and CBS). Moreover, whereas several of NBC's highest-rated shows in 1995 were in their first year of airing, the successful ABC shows were veterans. For example, one of ABC's highest rated shows is *Monday Night Football*, which was in its 25th season.

[68] In the estimation, $\varsigma^m_{i,j,t} = \upsilon^m_{i,j} + \varsigma^m_{New}New_{j,t} + \varsigma^m_{Veteran}Veteran_{j,t}$. The parameters ς^m_{New} and $\varsigma^m_{Veteran}$, which are supposed to capture the variation in consumers' familiarity with shows based on their veteran status and success, have the expected signs. However, their values are very small ($\varsigma^m_{New} = -0.004$ and $\varsigma^m_{Veteran} = 0.017$), and they are not different from zero even at the 10% significance level. This might result from the flexible form of $\upsilon^m_{i,j}$ that allows familiarity to vary across networks and individuals.

TABLE 4 The Effect of Advertising Exposures on λ

	Number of Exposures to Advertisements				
λ	0	1	2	3	4
ABC	0.720	0.793	0.834	0.861	0.880
CBS	0.442	0.763	0.848	0.888	0.912
NBC	0.653	0.789	0.847	0.880	0.901
Fox	0.404	0.808	0.886	0.918	0.937
Average	0.555	0.788	0.854	0.887	0.907

Note: The numbers in the table are the average of the $\lambda_{i,j,t}$ across individuals and time, where $\lambda_{i,j,t}$ is a measure of how well informed the consumer is. The numbers are directly computed from the structural estimates of the precision parameters and the relevant variables (e.g., an indicator function for veteran shows).

this table for the specific segments (posted at www.tau.ac.il/~rroonn/Papers/Matchmaker.html) illustrates the vast heterogeneity in viewers' product-specific knowledge even without any advertising exposures (λ ranges from 0.85 to 0.265). Furthermore, as expected, the effect of an advertising signal is lower for viewers who are more informed. For example, for the fourth segment, λ_{Fox} increases from 0.265 to 0.764 with exposure to a single advertisement, whereas for the second segment, λ_{ABC} hardly changes with advertising exposures (from 0.79 to 0.82). Because advertising signals inform individuals about product attributes, each additional signal increases the informativeness level of the individual. Thus, the effect of the nth signal is smaller than the effect of the $(n-1)$th. For example, whereas the average λ drops from 0.555 to 0.788 with the first exposure, it drops to 0.854 with the second exposure.

Section 6 presents additional ways to assess the behavioral implications of informative advertising.

□ **Goodness of fit.** The fit of the model in each of the 60 time slots is tested using the χ^2 test in Heckman (1984), which applies to models with parameters estimated from microdata. Constructing a single χ^2 statistic to test the model is not computationally feasible.[69] The test statistic is a quadratic form of the difference between observed cell counts and expected cell counts using the model. We reject the null hypothesis that the model is correctly specified for only 6 of the 60 time slots using a 5% significance level.

□ **A robustness check.** In the model section, we have assumed that although the individual is uncertain about $u^{att}_{i,j,t}$, she knows the expected value and the variance of $\xi_{j,t}$ and $x_{j,t}$ for each multiproduct firm. As a result of this assumption, consumers in our model rely on the multiproduct firms' profiles when forming their expectations about specific products. Indeed, in an earlier study (Anand and Shachar, 2004), we have shown that this assumption and its implication are supported by the data on television-viewing choices. Still, in order to assess the robustness of our results, it is interesting to relax this assumption and reestimate our model. We do this by allowing individuals' prior distribution of a program to depend on the attributes of *all* TV programs rather than only those of that particular network. The results of this exercise demonstrate the stability of our estimates—we find that $\widehat{\varsigma^a}$ is 0.2441 (with a standard error of 0.0616) versus 0.2739 (with a standard error of 0.0741) for the original specification with network-specific priors. In other words, the finding about the informativeness of ads and accordingly their matching role is robust to this alternative specification of the prior distribution.

[69] The number of cells that fully partition the response vector space is 5^{60} (when ignoring the absence of Fox in the 10:00–11:00 p.m. time slots and the initial conditions). This test is also a special case in Andrews (1998).

TABLE 5 Advertising Deters Consumption Using the Structural Estimates

	Among Individuals Who Watch TV in Time Slot t, the Percent of Them Who Choose Alternative j, Where the Chosen Attribute Utility ($\xi_{i,C_{i,t},t}$) Is in the:					
	Lowest				Highest	Number of
$N^a_{i,j,t}$	1%	10%	25%	33%	33%	Observations
0	.011	.111	.270	.355	.323	1935
1	.0003	.023	.129	.184	.356	211
2	0	.043	.074	.237	.394	42
	χ^2 Statistics for All the Comparisons					
0 versus 1	2.190	14.395**	14.767**	16.450**	0.635	
0 versus 2	0.462	1.735	5.940*	1.624	0.639	
1 versus 2	0.013	0.532	0.884	0.510	0.140	

Note: The top panel of the table represents the distribution of the chosen attribute utility, where the attribute utility is calculated using the structural parameters and the relevant variables. The bottom panel includes the χ^2 statistics to test whether the distributions (across the different rows of the top table) are the same.

For all the observations in this table, the number of advertising exposures is the same at time t across all the three big networks (ABC, CBS, and NBC).

*Significance at 5% level.

**Significance at 1% level.

6. Applications

■ This section illustrates the normative and positive consequences of informative advertising based on the structural estimates. It starts by examining the consumption-deterring aspect of advertising and its matchmaking role. The positive consequences center around targeting strategies of firms.

□ **Matching.** Section 2 provides preliminary evidence that a consumer's response to advertising depends on her taste for the product and, specifically, that exposure to advertising improves the matching of consumers and products. Reexamining these implications using the structural estimates reinforces these findings, as described below.

Consumption-deterring role. The model implies that exposures to advertising decrease the tendency to purchase products that do not fit consumer tastes well. Table 5 demonstrates that this implication is supported by our data.[70] It compares the tendency of consumers who were exposed to different numbers of ads to choose shows that yield low attribute utilities. Specifically, the columns in Table 5 differ with respect to the values of the *chosen* attribute utility for each individual. (As in previous sections, we focus on consumers who were exposed to the same number of ads to each of the competing networks in a specific time slot.) The structural estimates enable us to identify the values of attribute utilites $u^{att}_{i,j,t}$ for each combination of individual and show, and the table is based on the distribution of chosen attribute utilities.[71]

[70] Furthermore, in a working paper version of this study (Anand and Shachar, 2001), we demonstrate the consumption-deterring role of advertising using a nonstructural test based on viewing choices of news magazines. We focus on this distinct show genre because it is easy to distinguish between viewers who like this type of show and those who dislike it based on their viewing habits. We find that the tendency to watch a news magazine is increasing in ad exposures for consumers who like this type of show and decreasing in ad exposures for consumers who dislike this type of show.

[71] The task is complicated by the variation of tastes across the unobserved segments. This means that the attribute utility of individual i with show j might be low if the individual is of segment k but high if the individual is of segment k'. Thus, the unit of analysis in Table 5 is not the individual but rather the individual-segment combination, and instead of counting individuals, we "count" each individual six times (as the number of segments) and assign the segments'

The first column focuses on the lowest percentile of this distribution, the second column on the lowest decile, the third on the lowest quartile, and the fourth on the lowest 33rd percentile. As predicted by the model, the tendency to watch shows that yield low attribute utilities generally decreases in the number of ad exposures. As an example, consider the third column, which focuses on shows that yielded attribute utility in the lowest quartile. It shows that 27% of consumers who were not exposed to ads watched such shows, compared to 13% of those who were exposed to one ad and 7.4% of those who were exposed to two ads. By comparison, the last column presents the upper 33rd percentile of the distribution. There, the tendency of consumers to watch shows that lead to high utility increases in the number of ad exposures.[72]

Detecting the consumption-deterring aspect of advertising requires precise estimates of consumer preferences. This requirement makes a nonstructural examination difficult. Using the structural estimates, Table 5 reveals the strength of the consumption-deterring aspect in the data. Notice that the choices and the exposures to advertising in this table are based on the *actual* choices, not on predictions.

Matching of consumers with products. Advertising improves the match of consumers and products by guiding consumers to products that better fit their tastes. Next, we assess the magnitude of this matching effect in terms of the changes in both consumers' choices and their utility. The assessment is based on simulating consumers' choices under two scenarios for the number of exposures. In the first, the number of advertising exposures is equal to its value in the data set, $N^a_{i,j,t}$, and in the second it is equal to $N^a_{i,j,t} + 1$. The columns of Table 6 present these two scenarios.

The first panel of Table 6 describes the fit between individuals and the alternatives they chose for all the time slots. The fit depends, of course, on the utility that the individual derives from each alternative, and the utility is a function of various unobservables such as εs. In order to be able to determine the "quality" of consumer choices, we use simulated data for both scenarios.[73] Accordingly, whereas individuals do not know for certain which is their best alternative, we do. It turns out that although individuals are uncertain about product attributes, they choose their best alternative in about 80% of the cases for all the exposure scenarios in the first panel. The reason for this high percentage is the large magnitude of the state-dependence parameters. The increase in the fit between consumers and products as a result of an exposure to an additional advertisement is small when focusing on the percent of first-best choices (from 80.2% to 82.0%) but large when viewed from the perspective of non-first-best choices (a decrease from 19.8% to 18.0%).[74] This means that the number of cases in which individuals do *not* choose their first-best alternative decreases by about 9.1% (or 1.8 percentage points).

The value of informative advertising can be better assessed by focusing on times when individuals depart from the status quo, that is, when they switch. The second panel serves this purpose, by presenting the fit only for observations where the television was turned off in period $t - 1$ but was on in period t. Removing the effect of state dependence expresses itself in the smaller percent of first-best choices compared to the first panel. Furthermore, the effect of advertising is also larger than in the first panel. Specifically, whereas the percent of first-best choices is 42.6 in

posterior probabilities to these "counts." As a result, the number of observations in each cell is not an integer (e.g., the number of observations in the top left cell is $.011 * 1935 = 21.285$).

[72] The *only* comparison in the table that is not consistent with the hypothesis is between viewers who were exposed to two ads and those who were exposed to one ad for the case of the lowest 33rd percentile. The table also presents the χ^2 statistic for each of the possible comparisons. The difference between the first and the second rows (i.e., no exposure to ads versus one exposure) is statistically significant at the 1% significance level for three out of the five comparisons. The differences between the first two rows and the last one are usually not statistically significant. This is due to the small number of viewers who were exposed to two or more ads for each of the networks in the same time slot.

[73] The results are obtained using 100 simulation draws (of the product-specific signals and $\varepsilon_{i,j,t}$) for each individual and setting the persuasive effect to be equal to $g_i(N^a_{i,j,t})$ for both scenarios.

[74] All these changes are statistically significant at the 1% level.

TABLE 6 The Matching Effect of Advertising

		$N^a_{i,j,t}$	$N^a_{i,j,t}+1$
	All Time Slots (%)		
Chosen Alternative	First best	.802	.820
	Second best	.166	.160
	Third best	.025	.016
	Fourth best	.006	.002
$\bar{U}$		0.3614	0.4758
t value			101.8
	Time Slots in Which the Individual Just Turned the TV on (%)		
Chosen Alternative	First best	0.426	0.484
	Second best	0.315	0.326
	Third best	0.159	0.137
	Fourth best	0.073	0.042
$\bar{U}$		−2.4442	−1.9663
t value			156.8

Note: Using the structural parameters, observed variables, and random draws of the unobservables, we identify the utility of the individual with each show in any time slot and accordingly can rank the options based on their "quality."

The two columns: This table presents the distribution of choices over these four options under two scenarios: in the first, the number of advertising exposures is equal to its value in the data set, $N^a_{i,j,t}$, and in the second it is equal to $N^a_{i,j,t}+1$.

The last two rows in each panel: The row titled, $\bar{U}$ presents the expected utility given the distribution of choices, and the t value is for the hypothesis that $\bar{U}$ is the same in both scenarios.

The two panels: The first panel describes the fit between individuals and the alternatives they chose for all the time slots. The second panel focuses on the time slots in which the individual turned on the TV.

our data set, it is 48.4 for the scenario $N^a_{i,j,t}+1$. This also means that the percent of non-first-best choices falls from 57.4 to 51.6, a decrease of 5.8 percentage points or 10.1%.

Both panels include the average utility experienced by individuals, $\overline{U} \equiv \frac{1}{IT}\sum_{i=1}^{I}\sum_{t=1}^{T}\sum_{j=0}^{J}[U_{i,j,t}I\{\widehat{y}_{i,t}=j\}]$, where $\widehat{y}_{i,t}$ are the simulated choices. As discussed in the preliminary evidence (Section 2) and in the model section (Section 3), one would expect this realized utility to increase with the number of advertising exposures $N^a_{i,j,t}$. The results in Table 6 support this view. The change in utility is 0.1144 in the first panel and 0.4479 in the second panel. Because $\gamma_{Basic}=0.27$ per time slot, an increase in utility of 0.1144 as a result of an additional exposure to one advertisement for each show equals 42.3% of the increase in utility from having a cable connection.

□ **Targeting strategies and an informal specification test.** Advertisers face increasingly segmented audiences.[75] Segmentation provides firms an opportunity to improve the effectiveness of advertisements as long as they can identify the segment whose response to their advertisements is the strongest. The model presented here might assist in this task. By offering a new approach to model the effect of advertising (through both the utility and the information set), it provides a more precise estimate of the advertising elasticities for each consumer.

This subsection evaluates the benefits of this approach in targeting segmented audiences. This is done, first, by comparing the targeting strategies employed by the television networks with strategies suggested by our model as being "optimal." Then, the actual placements of advertisements are compared with the optimal ones. We find that the targeting strategies executed by the television networks are consistent with those implied by our model.

[75] In recent years, there has been a surge in the number of media outlets such as newspapers, television channels, magazines, websites, and so forth. For example, between 1985 and 1995, the number of significant cable channels increased from 31 to 87 (Parsons and Frieden, 1998). Similarly, between 1988 and 2001, the number of magazines wholly devoted to particular areas of interest such as "nursing" and "fish and fisheries" grew from 16 and 19, respectively, to 135 and 64 (www.magazine.org/resources/fact_sheets/ed7_8_01.html).

TABLE 7 Actual and Best Response Strategies

		Actual		Best Response	
Variable	Definition	Estimate	Standard Error	Estimate	Standard Error
Demographic_Match$_{j,k}$	The number of matches between the demographic characteristics of both shows. The demographic characteristics are age, gender, family status, and race	0.220	0.101	0.092	0.101
Genre_Match$_{j,k}$	A binary variable that is equal to one if the shows are from the same genre, and zero otherwise.	0.648	0.192	0.551	0.207
Preceding_Show$_{j,k}$	A binary variable that is equal to one if show k directly precedes show j, and zero otherwise.	1.787	0.354	2.546	0.422
Same_Hour$_{j,k}$	A binary variable that is equal to one if shows k and j are broadcast in the same hour, and zero otherwise.	0.379	0.186	0.383	0.197
Rating$_k$	The number of people who watch show k in our data set.	−0.200	3.187	26.452	3.363
Constant	—	−2.010	0.379	−3.816	0.406
McFadden R^2		0.0688		0.1407	

Note: The dependent variables in these logit estimations are as follows: for the columns labeled "actual," it is a binary variable that is equal to one if an advertisement for show j appeared in show k, and zero otherwise. For the "best response" columns, it is a binary variable that is equal to one if it is "optimal" (based on our structural estimates) to place an advertisement for show j in show k, and zero otherwise.

This exercise and its results can also be viewed as a nonformal specification test of the model. The reason is that although network executives never had access to the individual-level data that we employ, they have strong incentives to optimize their targeting strategies. Thus, the similarity between the predicted optimal behavior and the actual behavior suggests some confidence in the specification of the model.

In order to get a sensible comparison between the actual and the optimal placements, we compute the optimal strategies subject to two constraints. These constraints follow particular decisions made by the networks. First, the number of advertisements for each show in the exercise is set to be equal to the actual number. Second, the maximum number of advertisements for show j in show k (other than in a show that lasted two hours) is set to be one.

The logit models in Table 7 present the first comparison between actual and "optimal" strategies.[76] There are two dependent variables, $d^a_{j,k}$ and $d^p_{j,k}$. The binary variable $d^a_{j,k} = 1$ if an advertisement for show j appeared in show k, and is equal to zero otherwise. The binary variable

[76] The mechanism by which the optimal placement is chosen is as follows. (i) An advertisement for a show is hypothetically removed from the schedule, and replaced by an advertisement for the same show in the first time slot of the week. The total predicted rating of the network is then calculated. (ii) Step 1 is repeated, this time by placing the advertisement in the second time slot, and then in every subsequent one until the time slot that precedes the show in question. (iii) The advertisement in question is placed in the time slot that yields the highest predicted rating. (iv) Steps i–iii are repeated for every advertisement in the schedule, holding the previous advertisements in their optimal locations. (v) This entire cycle (steps i–iv) is repeated until no ad location changes through an entire cycle. Another mechanism to calculate the optimal placements is to change the locations of advertisements for all the shows simultaneously rather than sequentially. However, because the computation of this alternative mechanism is infeasibly complex, we employ the mechanism described above. Obviously, this means that the chosen mechanism only provides a lower bound on rating improvements. Although the network executive maximizes profits, we ignore cost considerations and focus on ratings instead of revenues. Goettler (1999) shows that this approximation is not costly when computing an optimal show schedule.

$d^p_{j,k} = 1$ if it is optimal (based on the model) to place an advertisement for show j in show k, and zero otherwise.

The first four independent variables in the table are measures of match between shows j and k ($Demographic\_Match_{j,k}$, $Genre\_Match_{j,k}$, $Same\_Hour_{j,k}$, and $Preceding\_Show_{j,k}$), and the last variable ($Rating_k$) measures the ratings of show k. All these variables are formally defined in the table. The variables $Demographic\_Match_{j,k}$ and $Genre\_Match_{j,k}$ are based on the similarity in the attributes of shows j and k. The binary variables $Same\_Hour_{j,k}$ and $Preceding\_Show_{j,k}$ are equal to one if shows j and k are broadcast in the same hour of a night, and if show k directly precedes show j, respectively. Last, $Rating_k$ is equal to the number of people who watched show k in our data set. The independent variables are not based on theory but rather on the model estimates and intuition. One should expect the variables $Demographic\_Match_{j,k}$ and $Genre\_Match_{j,k}$ to have a positive effect on the probabilities of the events $d^a_{j,k} = 1$ and $d^p_{j,k} = 1$ because of the informative role of advertising. Specifically, the model implies that a consumer's response to advertising is higher when the product better fits her tastes. Accordingly, the coefficient on $Same\_Hour_{j,k}$ is expected to be positive because individuals tend to watch television in regular time slots. A similar argument based on state dependence suggests that the coefficient on $Preceding\_Show_{j,k}$ should be positive. The expected sign of the coefficient on $Rating_k$ is positive when the dependent variable is $d^p_{j,k}$ because show k provides a larger number of individuals who would be exposed to the advertisement. When the dependent variable is $d^a_{j,k}$, however, the expected sign is ambiguous because larger ratings also imply higher opportunity costs.

In general, the coefficients have the expected signs and the actual strategies are consistent with the optimal ones. Specifically, the effect of the genre match variable on both $d^p_{j,k}$ and $d^a_{j,k}$ is positive and different from zero at the 1% significance level. Also, the effect of the demographic match on both $d^p_{j,k}$ and $d^a_{j,k}$ is positive.[77] These findings suggest that it is optimal to place advertisements for show j in show k when the two shows have similar attributes. Furthermore, the coefficients are quite similar in both the logit models (with $d^a_{j,k}$ and $d^p_{j,k}$ as dependent variables), implying that the networks usually do follow such a strategy. The coefficients of $Preceding\_Show_{j,k}$ and $Same\_Hour_{j,k}$ are, as expected, positive in both models. This means that it is optimal to place advertisements in the preceding show and in the same hour (in previous nights) and that the networks apply this strategy. As expected, there is a difference in the effect of ratings ($Rating_k$) between the two models.

Another way to evaluate how close the actual strategies are to the optimal ones is by directly comparing the placements themselves. The percent of advertisements that are placed in the optimal locations is 64%. Furthermore, the model predicts that if the networks were to replace their locations with the optimal ones, their ratings would have increased by only 4.4%. (Specifically, the market share of ABC would have increased from .0777 to .0811, of CBS from .0565 to .0602, of NBC from .0874 to .0901, and of Fox from .0487 to .0505.) Despite that our exercise does not account for the cost of placing advertisements, these results indicate that the network strategies are quite close to the ones that the model should predict as optimal.[78]

We have also solved for a Nash equilibrium and found that the percent of advertisements that are placed in the optimal locations is 62%.[79] The model predicts that if the networks were to replace their locations with the equilibrium placements, their ratings would have increased by only 3.3%. (Specifically, the market share of ABC would have increased from .0777 to .0801, of CBS from .0565 to .0595, of NBC from .0874 to .0891, and of Fox from .0487 to .0500.)

[77] However, although the coefficient for the actual behavior is different from zero at the 1% level, it is not possible to reject the hypothesis that the coefficient for the optimal behavior is zero even at the 10% significance level.

[78] Recall as well that optimal strategies are constrained to follow certain decisions made by the networks. When these constraints are released, the rating gains are higher than those reported in the text.

[79] For any network A, its best response is calculated as described above. Given this new schedule of advertising for A and the actual schedule for C and D, the best response of network B is now computed. The process is repeated until a full cycle of the four networks does not yield any change in the schedule.

7. Conclusion

■ The findings in this study are relevant for both theoretical and empirical work. Here we highlight potential extensions and applications of these findings.

A vast literature in economics studies how firms strategically reveal information to consumers through signalling. In practice, consumers do not obtain information only by inference from firm behavior. This study shows that consumers rely on advertising content to directly obtain information about product attributes and, as a result, improve their match with products. This matching role of advertising is likely to be increasingly important as product offerings continue to proliferate in many markets and consumers find it hard to remain informed of product attributes. In revealing the matching role of advertising, the model and data presented here allow us to address two central empirical challenges that confront advertising studies: first, identifying the informative effect of advertising separately from its direct effect on the utility; and, second, the endogeneity of advertising exposures.

Advertising is not the only way, of course, in which firms inform consumers. Automobiles produced by Volvo, for example, are perceived by consumers to be safe, and movies produced by Disney are perceived by viewers to be family friendly. These multiproduct firms and others provide information by selecting a clear product line. Elsewhere we show that the information set of consumers indeed depends on the profile of multiproduct firms (Anand and Shachar, 2004). Together, these findings raise several questions that merit further research. For example, what does theory imply about the relationship between advertising intensity and product diversity in equilibrium? And, what are the consequences of the informational role of both advertising and product line choices for consumer welfare?

An increased demand for segmented media channels, coupled with technological advances, has created new services (e.g., TiVo) that enable firms to target consumers individually. The theoretical and empirical consequences of informative advertising and individualized targeting services merit further attention. We draw attention to two aspects here. First, the finding in this study that advertising can deter consumption raises interesting issues concerning the targeting strategies of firms. For example, it implies that poor targeting need not be payoff neutral, but can even reduce a firm's market share. The danger of exposing the wrong consumer to advertisements creates sharp incentives for precise targeting and increases the demand for media channels that deliver highly segmented audiences. Second, consumers can also learn about product attributes from firms' targeting strategies themselves. The equilibrium of markets with rational consumers, informative advertising, and segmented media channels is the focus of Anand and Shachar (2009). It is shown there that, in equilibrium, rational consumers respond positively to advertising intensity. One interesting implication of this is that the effect of advertising intensity on the utility (captured in the model here by ρ) may be a consequence of rational behavior.

Relatedly, the growing importance of advertising has led researchers to include advertising intensity as an independent variable in their demand estimation. To the extent that such models do not allow the information set to depend on advertising *content*, the findings in this article suggest that they might, as a result, be misspecified. Of course, suggesting that each of these models employs exactly the same approach taken here is unreasonable due to the complexity of such a task. This raises the need to develop a simple modelling approach that can proxy the effect of advertising through the information set. For example, one approach is to allow ρ to be a function of consumer tastes. Furthermore, although the evidence presented in Sections 2 and 5 illustrates simple ways to assess the informative role of advertising, additional uncomplicated examinations can be fruitful. For example, many firms allow consumers to return products after purchase. Data on exposures to advertising combined with information on product returns can be used to reveal the informative role of advertising.

Finally, although we have outlined the advantages of the data set created for this study, such data are likely to have a caveat. Specifically, advertisements for television shows can be thought of as product samples and as a result may be more informative than advertising for other products. Examining the robustness of our results would require creating similar data sets for other products.

References

ACKERBERG, D. "Empirically Distinguishing Informative and Prestige Effects of Advertising." *RAND Journal of Economics,* Vol. 32 (2001), pp. 316–333.

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

ANAND, B. AND SHACHAR, R. "Advertising, the Matchmaker." www.tau.ac.il/~rroonn/Papers/match2001.pdf, 2001.

—— AND ——. "Brands as Beacons: A New Source of Loyalty to Multiproduct Firms." *Journal of Marketing Research,* Vol. 41 (2004), pp. 135–150.

—— AND ——. "Targeted Advertising as a Signal." *Quantitative Marketing and Economics,* Vol. 7 (2009), pp. 237–266.

ANDERSON, S. AND RENAULT, R. "Advertising Content." *American Economic Review*, Vol. 96 (2006), pp. 93–113.

ANDREWS, D. "Chi-Square Diagnostic Tests for Econometric Models: Introduction and Applications." *Journal of Econometrics,* Vol. 37 (1998), pp. 135–156.

ARELLANO, M. *Panel Data Econometrics*. Oxford: Oxford University Press, 2003.

BAGWELL, K. "The Economic Analysis of Advertising." In M. Armstrong and R. Porter, eds., *Handbook of Industrial Organization*. Amsterdam: North-Holland, 2003.

BECKER, G. AND MURPHY, K. "A Simple Theory of Advertising as a Good or Bad." *Quarterly Journal of Economics,* Vol. 108 (1993), pp. 941–964.

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

——, LEVINSOHN, J., AND PAKES, A."Automobile Prices in Market Equilibrium." *Econometrica,* Vol. 63 (1995), pp. 841–890.

BOND, J. AND KIRSHENBAUM, R. *Under the Radar: Talking to Today's Cynical Consumer*. New York: John Wiley & Sons, 1998.

BUTTERS, G. "Equilibrium Distribution of Prices and Advertising." *Review of Economic Studies,* Vol. 44 (1977), pp. 465–492.

CARRASCO, R. "Binary Choice with Binary Endogenous Regressors in Panel Data: Estimating the Effect of Fertility on Female Labour Participation." *Journal of Business and Economic Statistics,* Vol. 19 (2001), pp. 385–394.

CHAMBERLAIN, G. "Multivariate Regression Models for Panel Data." In G.S. MADDALA, ed., *The Econometrics of Panel Data,* Vol. 1. Aldershot, UK: Edward Elgar Publishing, 1993.

CHERNEW, M., GOWRISANKARAN, G., AND SCANLON, D.P. "Learning and the Value of Information: The Case of Health Plan Report Cards." *Journal of Econometrics,* Vol. 144 (2008), pp. 156–174.

CHING, A. "Dynamic Equilibrium in the U.S. Prescription Drug Market after Patent Expiration." Contributed Paper no. 1242, Econometric Society World Congress 2000, 2000.

COSCELLI, A. AND SHUM, M. "An Empirical Model of Learning and Spillovers in New Drug Entry." *Journal of Econometrics,* Vol. 122 (2004), pp. 213–246.

CRAWFORD, G. AND SHUM, M. "Uncertainty and Learning in Pharmaceutical Demand." *Econometrica,* Vol. 73 (2005), pp. 1137–1173.

DEGROOT, M. *Probability and Statistics,* 2d ed. Boston: Addison Wesley, 1989.

ECKSTEIN, Z., HORSKY, D., AND RABAN, Y. "An Empirical Dynamic Model of Optimal Brand Choice." Working Paper no. 88, Foerder Institute of Economic Research, 1988.

ERDEM, T. AND KEANE, M. "Decision-Making under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets." *Marketing Science,* Vol. 15 (1996), pp. 1–20.

GOETTLER, R. "Advertising Rates, Audience Composition, and Competition in the Network Television Industry." GSIA Working Paper no. 1999-E28, Department of Economics, Carnegie Mellon University, 1999.

—— AND SHACHAR, R. "Spatial Competition in the Network Television Industry." *RAND Journal of Economics,* Vol. 32 (2001), pp. 624–656.

GROSSMAN, G. AND SHAPIRO, C. "Informative Advertising with Differentiated Products." *Review of Economic Studies,* Vol. 51 (1984), pp. 63–82.

HAJIVASSILIOU, V. "Some Practical Issues in Maximum Simulated Likelihood." Discussion Papers Series no. EM/97/340, London School of Economics, 1997.

—— AND RUUD, P. "Classical Estimation Methods for LDV Models Using Simulation." In R.F. Engel and D.L. McFadden, eds., *Handbook of Econometrics,* Vol. IV. Amsterdam: Elsevier Science, 1994.

HECKMAN, J. "Statistical Models for Discrete Panel Data." In C. Manski and D. McFadden, eds., *Structural Analysis of Discrete Data with Econometric Applications*. Cambridge, Mass.: MIT Press, 1981a.

——. "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time–Discrete Data Stochastic Process." In C. Manski and D. McFadden, eds., *Structural Analysis of Discrete Data with Econometric Applications*. Cambridge, Mass.: MIT Press, 1981b.

——. "The χ^2 Goodness of Fit Statistic for Models with Parameters Estimated from Microdata." *Econometrica,* Vol. 52 (1984), pp. 1543–1548.

JACOBY, J. AND HOYER, W. "Viewer Miscomprehension of Televised Communication: Selected Findings." *Journal of Marketing,* Vol. 46 (1982), pp. 12–26.

—— AND ——. "The Comprehension/Miscomprehension of Print Communication: Selected Findings." *Journal of Consumer Research,* Vol. 15 (1989), pp. 434–443.

JIN, G. AND LESLIE, P. "The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards." *Quarterly Journal of Economics,* Vol. 118 (2003), pp. 409–451.

JOHNSON, J. AND MYATT, D. "On the Simple Economics of Advertising, Marketing and Product Design." *American Economic Review,* Vol. 96 (2006), pp. 756–784.

KIHLSTROM, R. AND RIORDAN, M. "Advertising as a Signal." *Journal of Political Economy,* Vol. 92 (1984), pp. 427–450.

KRUGMAN, H.E. "The Learning of Consumer Likes, Preferences and Choices." In F.M. Bass, C.W. King, and E.A. Pessemier, eds. *Applications of the Sciences in Marketing Management*. New York: Wiley, 1968.

LANCASTER, K. *Consumer Demand: A New Approach*. New York: Columbia University Press, 1971.

MAHAJAN, V., MULLER, E., AND WIND, J. *New Product Diffusion Models*. Boston: Kluwer Academic, 2000.

MANKIW, G. *Principles of Economics*. New York: Dryden Press, 1998.

MCFADDEN, D. "Conditional Logit Analysis of Qualitative Choice Behavior." In P. Zarembka, ed., *Frontiers in Econometrics*. New York: Academic Press, 1973.

——. "Econometric Models of Probabilistic Choice." In C. Manski and D. McFadden, eds., *Structural Analysis of Discrete Data with Econometric Applications*. Cambridge, Mass.: MIT Press, 1981.

——. "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration." *Econometrica,* Vol. 57 (1989), pp. 995–1026.

MILGROM, P. AND ROBERTS, J. "Price and Advertising Signals of Product Quality." *Journal of Political Economy,* Vol. 94 (1986), pp. 796–821.

MILYO, J. AND WALDFOGEL, J. "The Effect of Price Advertising on Prices: Evidence in the Wake of 44 Liquormart." *American Economic Review,* Vol. 89 (1999), pp. 1081–1096.

National Reference Supplement. Nielsen Media Research, 1995–1996.

NELSON, P. "Advertising as Information." *Journal of Political Economy,* Vol. 82 (1974), pp. 729–734.

NEVO, A. "Measuring Market Power in the Ready-to-Eat Cereal Industry." *Econometrica,* Vol. 69 (2001), pp. 307–342.

PAKES, A. AND POLLARD, D. "Simulation and the Asymptotics of Optimization Estimators." *Econometrica,* Vol. 57 (1989), pp. 1027–1057.

PARSONS, P. AND FRIEDEN, R. *The Cable and Satellite Television Industries*. Boston: Allyn and Bacon, 1998.

ROBERTS, M.J. AND SAMUELSON, L. "An Empirical Analysis of Dynamic, Nonprice Competition in an Oligopolistic Industry." *RAND Journal of Economics,* Vol. 19 (1988), pp. 200–220.

RUBINSTEIN, R. *Simulation and the Monte Carlo Method*. New York: John Wiley & Sons, 1981.

RUST, R. AND ALPERT, M. "An Audience Flow Model of Television Viewing Choice." *Marketing Science,* Vol. 3 (1984), pp. 113–124.

SHACHAR, R. "A Diagnostic Test for the Sources of Persistence in Individuals' Decisions." *Economics Letters,* Vol. 45 (1994), pp. 7–13.

—— AND ANAND, B. "The Effectiveness and Targeting of Television Advertising." *Journal of Economics and Management Strategy,* Vol. 7 (1998), pp. 363–396.

—— AND EMERSON, J. "Cast Demographics, Unobserved Segments, and Heterogeneous Switching Costs in a TV Viewing Choice Model." *Journal of Marketing Research,* Vol. 37 (2000), pp. 173–186.

STERN, S. AND TRAJTENBERG, M. "Empirical Implications of Physician Authority in Pharmaceutical Decisionmaking." Working Paper 6851, 2001.

U.S. BUREAU OF THE CENSUS, *Statistical Abstract of the United States* (1995), Section 14, no. 728. Washington, Government Printing office.

ZAJONC, R. "Attitudinal Effects of Mere Exposure." *Journal of Personality and Social Psychology,* Vol. 9 (1968), pp. 1–28.