YITING DENG and CARL F. MELA*

Television, the predominant advertising medium, is being transformed by the microtargeting capabilities of set-top boxes (STBs). By procuring impressions at the STB level (often denoted "programmatic television"), advertisers can now lower per-exposure costs and/or reach viewers most responsive to advertising creatives. Accordingly, this study uses a proprietary, household-level, single-source data set to develop an instantaneous show and advertisement viewing model to forecast consumers' exposure to advertising and the downstream consequences for impressions and sales. Viewing data suggest that person-specific factors dwarf brand- or show-specific factors in explaining advertising avoidance, thereby suggesting that device-level advertising targeting can be more effective than existing show-level targeting. Consistent with this observation, the model indicates that microtargeting lowers advertising costs and raises incremental profits considerably relative to show-level targeting. Further, these advantages are amplified when advertisers can buy in real time as opposed to up front.

Keywords: TV advertising, targeting, set-top box, sampling, programmatic television

Online Supplement: http://dx.doi.org/10.1509/jmr.15.0421

TV Viewing and Advertising Targeting

Television is the most prominent modality for the transmission and reception of video content. According to Nielsen's total audience report (Nielsen 2016), American adults spend about 4 hours and 39 minutes watching traditional TV each day, far exceeding the combined daily consumption of PCs, smartphones, and tablets. Despite fierce competition from digital advertising, TV advertising is still huge and growing. For example, eMarketer forecasts that TV spending will be $71.29 billion by the end of 2016 (eMarketer 2016). One of the key advantages of Internet advertising over TV has been its ability to target households. However, with the proliferation of digital set-top boxes (STBs) and digital video recorders (DVRs) that enable advertisers to target households, this advantage is eroding. It is estimated that the number of U.S. households reachable by addressable TV ads is now 68 million. Annual revenue for so-called programmatic TV advertising was $300 million in 2015, with an astonishing 100% year-over-year growth in revenue (Urbanski 2016), and the *Wall Street Journal* reports the market will soon reach $100 billion (Bruell 2017).

The digital targeting of advertising on television is transformative in two regards. First, digital TV affords household-level measures of TV viewing, making it possible to better forecast household advertising exposure. Second, STBs enable targeting at the household-show level, enhancing targeting precision. While improving the cost efficiency of advertising exposures is of interest in its own right, coupling set-top viewing data with purchase data can also be used to evaluate the profitability of ad campaigns at the household-show level.

As a result, several companies have already started offering targeted TV advertising solutions. For instance, TiVo Research and Analytics (TRA) developed a software platform, Media TRAnalytics that combines household-level TV tuning and purchase data to help advertisers achieve higher return on investment (ROI). Similarly, Nielsen Catalina Solutions launched

*Yiting Deng is Assistant Professor of Marketing, UCL School of Management, University College London (email: yiting.deng@ucl.ac.uk). Carl F. Mela is T. Austin Finch Foundation Professor of Business Administration, Fuqua School of Business, Duke University (email: mela@duke.edu). A portion of this work was completed while the second author was visiting at the Rotterdam School of Management, Erasmus University. The authors would like to thank Bryan Bollinger, Pedro Gardete, Ron Shachar, Richard Staelin, Peng Sun, Kenneth Wilbur, Yi (Daniel) Xu, participants at the 2016 Marketing Science Conference, and seminar participants at Duke University, University of California–Riverside, Erasmus University, HKUST, Rotterdam School of Management, McGill University, New York University, University of Miami, University of Rochester, University of Notre Dame, Syracuse University, Cornell University, SUNY Buffalo, Peking University, and University College London, for useful comments. All errors are our own. Coeditor: Rajdeep Grewal; Associate Editor: Gerard Tellis.

ISSN: 0022-2437 (print)
1547-7193 (electronic)

Journal of Marketing Research
Vol. LV (February 2018), 99–118
DOI: 10.1509/jmr.15.0421

AdVantics on Demand, which helps advertisers achieve better targeting based on retail purchase data, and NBCUniversal announced the launch of its Audience Targeting Platform (ATP), which will use viewing and purchase histories to identify client-specific inventory across NBCUniversal's portfolio of national broadcast and cable networks. Acxiom TV, launched in January 2016, enables advertisers to combine first- and third-party data with TV viewing information to target advertising on TV. In a venture called Open AP, Viacom, 21st Century Fox, and Time Warner are coordinating to standardize purchasing of programmatic advertising.

In response to these evolving advances in digital TV, our goal is to use set-top data to model households' instantaneous TV and advertising consumption behavior and integrate the model with purchase data to propose ways for advertisers to improve their targeting profitability. We consider several questions. First, how do viewers[1] make TV and advertising consumption decisions, and are viewing data informative about these choices? Second, are set-top advertising exposures informative about purchase behaviors? Last, what are the implications of these advertising and purchase consumption decisions for microtargeting advertisements on TV? To answer these questions, we combine panel data on TV viewing, consumer purchase-behavior panels, advertising campaign information, advertising rates, TV program descriptors, TV program ratings, and advertising rates.

With regard to TV and advertising consumption, we document a number of stylized facts about viewing to inform our modeling choices. For example, 85% of households watch prime-time TV every night, on average watching 88% of the prime-time hours. Thus, any gains from targeting largely arise from what a viewer watches rather than whether they watch. Once a viewing session commences, 30% of households sample multiple shows before selecting one to view, suggesting that show sampling is informative about viewing preferences. Once a show is selected, 60% of viewers watch a show to its conclusion, indicating ad exposure is common once a show is selected. Within a show, we find that viewers' advertising avoidance is more common when the show is recorded (79% vs 15%). Person-specific factors explain most of the variation in advertising avoidance (20.4%). The person-specific variation dwarfs the variation explained by genre (.7%), brand (.3%), time (.0%), and category (.0%). This variance decomposition suggests that there is potential for significant returns to household-level advertising targeting and that targeting on genres and time will have smaller effects on advertising avoidance.

Using these insights to develop an integrated model of viewing predicated on a single consumption utility framework, we capture the show sampling, show viewing, and show exit processes described above. The approach builds on Arcidiacono et al.'s (2015) continuous-time dynamic discrete-choice models, which we extend by incorporating sampling/consideration behavior. Factors playing a major role in show consumption utility include a show's length, genre, network, familiarity, and offset. When consumers encounter an advertisement in these shows, we again draw upon this concept of consumption utility to predict whether consumers will avoid an advertisement. Results suggest advertising utility is lower when the prior advertisement is avoided, and in reality and drama genres; the utility is higher in the first slot of a commercial break.

[1]We use the terms "household" and "viewer" interchangeably.

Consistent with prior research (e.g., Sethuraman, Tellis, and Briesch 2011), we observe small but generally positive advertising effects on purchase behavior. Across 129 ad campaigns spanning multiple categories and consumer segments, we find 15% evidence positive advertising effects at a statistical significance level of 5%. Of note, we find variation across consumers in their responsiveness to advertising, suggesting that set-top data, coupled with purchase information, can be exploited to target households most responsive to advertising.

With regard to the microtargeting of advertisements, we consider (1) whether an advertiser seeks to minimize the costs of its target advertising exposures or to maximize its incremental profit from advertising and (2) whether the advertising purchase is made in advance or in real time. The advantage of cost-based targeting is that it is predicated upon only advertising viewership and does not require a model of sales response to advertising. However, cost-based targeting ignores the link between advertising and sales and thus, the profitability of a campaign. The advantage of real-time advertising is knowing what a viewer is watching, averting the need for a TV show viewing model. Purchasing in advance, or "up front," is the current norm in TV advertising sales, but with the increased potential for firms to buy advertising in real time, as on the Internet, real-time buying is becoming increasingly relevant. With advance buying, we find that it is possible to lower costs per target ad view by over 90%. With real-time buying, it is further possible to lower target costs per view while concurrently increasing target views; in one schedule, views to the target households can be increased by 47% while concurrently reducing costs by 7%. In short, we document dramatic increases in the cost efficiency of targeted media buys. Likewise, we find that advertisers can improve their advertising profitability when real-time ad buys are enabled. The effect is smaller for advance buys because uncertainty about advertising exposures attenuates advertising response, making advertising less profitable.

The remainder of the article is organized as follows. In the next section, we review the relevant literature. We then describe TV viewing behavior to motivate the viewing model presented in the following section. Subsequently, we discuss estimation and describe the estimation results. Based on these results and a purchase model, we conduct counterfactual policy experiments and evaluate potential gain to be realized from targeting. Finally, we conclude with a schedule of next steps.

RELEVANT LITERATURE

In this section, we detail how this research relates to the literature on television viewing, ad viewing, and ad targeting. Since Lehmann's (1971) seminal work, a rich body of literature has identified various factors affecting viewers' utility from watching TV programs. Such factors include viewer demographics (Anand and Shachar 2011; Rust and Alpert 1984), program genre (Anand and Shachar 2011; Rust and Alpert 1984), cast demographics (Shachar and Emerson 2000; Wilbur 2008), advertising time (Wilbur 2008), viewer's previous program choices (Moshkin and Shachar 2002), and spouse's choice (Yang, Narayan, and Assael 2006). Collectively, this line of literature suggests that person, show, and time factors explain substantial variation in-show viewing. Accordingly, we integrate these various factors into a household-level viewing model.

Conditional on watching TV shows, viewers inevitably encounter commercial breaks, which often trigger "zapping" (i.e., channel switching, leaving the room, etc.), and, in the case of recorded shows, "zipping" (i.e., fast-forwarding). Hence, a second related stream of literature examines viewers' advertising avoidance behavior and has identified various viewer- and ad-specific factors that affect such behavior. Identified viewer-specific factors include household category purchase history (Siddarth and Chattopadhyay 1998) and the media weight of a campaign (i.e., the number of times a household had previously been exposed to a commercial) (Gustafson and Siddarth 2007). Identified ad-specific factors include the frequency of the commercial (Siddarth and Chattopadhyay 1998), length and content of the commercial (Siddarth and Chattopadhyay 1998), program genre (Schweidel and Kent 2010), and commercial location (Gustafson and Siddarth 2007), as well as the congruity between the commercial and the program (Furnham and Price 2006).

Finally, given our goal to explore the potential of microtargeting in TV advertising, this work also relates to the growing literature in advertising targeting. Several studies have examined why and how targeted TV advertising works from a theoretical point of view (Anand and Shachar 2009; Gal-Or et al. 2006). A few studies address the issue of geographically or demographically targeted TV advertising from an empirical point of view (Anand and Shachar 2011; Lovett and Peress 2015). Our focus is instead targeting at more granular levels, namely, the individual household. Finer targeting affords better opportunities to incorporate individual households' past viewing and purchase data in targeting decisions. In this regard, the most closely related study is that of Tuchman, Nair, and Gardete (2017), who explore implications of individual-level targeting to consumers who are less likely to avoid the targeted advertisements and have a positive marginal advertising effect on purchase. One key point of departure in our analysis is that we model instantaneous show viewing, which enables us to forecast whether an advertisement is viewed at a given point in time. A second point of difference is that we consider not only which advertisements to target to whom, but in which show and at what time. This requires a model of show and advertising viewing behavior.

DATA

Microtargeting in TV is facilitated by historical viewing and purchase data at the household level. We first overview these sources of the household-level viewing and purchase data in the "Data Description" section. Next, we describe the TV program viewing data and the advertising viewing data to generate insights regarding household viewing behavior (in the "Empirical Regularities in TV Viewing" section) that help to form the basis of our viewing model.

Data Description

Several sets of data are integrated for this study, including (1) STB usage data, (2) purchase data, (3) advertising data, (4) programming data, and (5) Nielsen TV viewing data. The combination of these first four sets is often called single-source data because it covers the entire TV viewing and purchase experience for a set of households. The STB data (TiVo log files) track each household's complete usage of a TiVo STB and, therefore, all viewing behavior. The purchase data are from Information Resources Incorporated (IRI) and contain each household's store visits and purchase history in 77 consumer packaged goods (CPG) categories, as well as store causal data. The advertising data are obtained from TNS Media Research and include the timing and advertising costs for national TV advertisements airing within the duration of the data. These data are supplemented with national viewing data from Nielsen to normalize shows' advertising rates to the exposure level. The programming data come from Tribune Media Services (TMS) and contain information on popular TV programs. The STB data, advertising data, and programming data will be used to estimate the viewing model. The Nielsen TV viewing data contain rating information and advertising rates (costs). The purchase data and the Nielsen data will be used along with viewing model estimates in policy experiments on targeting.

TiVo log files (show and advertising viewing). The viewing data are from a field study conducted by IRI, TiVo, and a consortium of major CPG manufacturers (for a detailed description of the field study, see Bronnenberg, Dubé, and Mela 2010). The TiVo log files track each household's moment-by-moment usage of a TiVo STB. They record every keystroke of the STB as well as all TV content viewed and whether it was live or recorded. Among other things, the keystrokes are used to determine which content was fast-forwarded. We use data in the period July 2005–July 2006, keeping the 834 households for which we have both viewing and purchase information.

IRI data sets (purchase). The panel data used to link with TV viewing and advertising exposure data are provided by IRI and include purchase data, trip data, and store data in the period June 2005–June 2006.[2]

The first component, the IRI purchase panel data, contains the purchase history for panelists in 77 categories. Organized by panelist–category–item–transaction time, the data include store, item, item attributes, price, and promotional status (display or feature).

The second component, the IRI trip panel data, records panelists' store visits. Organized by panelist–transaction time, the data include the store visited and total amount spent. Combined with the purchase panel data, these store visit data enable us to infer nonpurchases, defined as no purchase in a category on a given store visit.

The third component, the IRI store causal data, reports store sales for each item sold in the 57 stores. Organized by store-week-item, the data include weekly price, promotional status (display or feature), and units sold. By matching these data with transactions in the purchase panel data and store visits in the trip panel data, we can construct a choice set with associated causal variables for each purchase occasion.

TNS advertising schedule data (advertising exposures, creative execution, and prices). The TNS advertising schedule data describe advertising schedules for 61 national broadcast and cable TV networks. For each advertisement, the data report the precise air time, network, length, attributes of the advertised product (e.g., product category, company, and brand), name and genre of the associated show, location of the commercial break within the show (i.e., pod) and the slot

[2]The starting and ending dates of the IRI data sets are both earlier than the respective dates for the three data sets related to TV viewing. All data sets intersect during the period July 2005–June 2006. We retain the IRI data in June 2005, one month before the start of the TV data, in order to initialize behavioral measures such as last brand purchased. We retain the TV data in July 2006, one month after the end of the IRI data, for holdout validation and policy experiments.

within the break (i.e., pod location or slot), and estimated price of the advertisement. We infer advertising exposures by noting the time and network of the advertisement and assessing whether the network was viewed at that time.

The data also contain a brief description of the advertisement, which is a summary of the creative execution for the advertisement (e.g., "mega roll/bear changes role" for Charmin bathroom tissue). A creative execution is often run for several weeks and is used to analyze advertising response. In the "Advertising Response" section, we will consider the effect on sales of dropping a particular advertising creative.

Nielsen TV viewing data. Because advertising rates furnished by TNS are at the show level, they do not yield a per-impression cost. Microtargeting is at the impression level, so we translate the show cost to an exposure cost by collecting information on advertising exposures. Specifically, we supplement the TNS advertising schedule data with Nielsen ratings to obtain per-exposure price for each advertisement. The Nielsen ratings are manually collected from *Broadcasting & Cable* magazine and report the audience size of the top TV programs on broadcast networks. We collect these data in July 2006, the period during which the policy experiments are run. These data include 376 shows on four networks: ABC, CBS, NBC, and FOX.

TMS data (program characteristics). TMS data contain descriptive information (e.g., program name, genre, cast, plot description) for 55,684 programs accounting for 90% of the TiVo viewing observations related to the top 27 TV networks (six broadcast networks and 21 cable networks). Each observation in the TiVo Log Files is tagged with a unique TMS identifier, which is used to match with the TMS data, resulting in a description of each show viewed.

Empirical Regularities in TV Viewing

This section reports a descriptive analysis of TV viewing both to illustrate the nature of the data and to motivate the ensuing model. Households' viewing behavior can be described by a series of conditional decisions, and we organize the discussion along this progression of conditional decisions (watch TV, sample shows, watch or record shows, watch advertisements, exit show/exit TV).

Watching TV. Owing to the observation that most viewing (and advertising spending) takes place in prime time (defined in our analysis as 8 P.M.–midnight local standard time), our ensuing analyses focus on this daypart. Figure 1, Panel A, shows most households watch TV most evenings: on average, households watch prime-time TV on 85% of the days in the sample. Panel B further depicts the hours of prime-time TV viewing across household-days, conditional on watching TV. Over 75% of the household-day viewing time exceeds 3 hours.

Show sampling and viewing. When starting to watch TV, a household first chooses which show to watch from among the live broadcast shows and the inventory of recorded shows available on the STB. Due to incomplete knowledge of program and episode quality, a household may sample shows (either live or recorded) for a brief duration to decide whether to watch (Esteves-Sorenson and Perretti 2012). After viewing the show for a short period of time, the household can decide whether to continue watching, sample another show, or exit. If another show is sampled, the process repeats.

To characterize the process of show sampling, we define "show completion rate" as the ratio of the time a show is viewed relative to its total broadcast length. Figure 1, Panel C, illustrates that completion rate is bimodal: it tends to be either very low or very high. This dichotomy is consistent with a process wherein people first sample a set of shows and then proceed to watch the one that is liked. Truncating the nonviews and the completed views (the end points in Figure 1, Panel C) enables us to zoom in on the sampling behavior (Figure 1, Panel D). It suggests a power law with most sampling not exceeding a few minutes per show.[3]

Further illustration of the apparent sampling process requires a definition of sampling events. Because excessively short viewing durations (e.g., 30 seconds or less) likely reflect channel "surfing" (i.e., using the "up" or "down" button to shift channels), a sampling duration is defined as 30 seconds or longer. To obtain the threshold that differentiates watching from sampling, we collect one-hour shows watched from the broadcast start time and compute the hazard rate (i.e., the fraction of surviving viewers that leave) by time into show (Figure 1, Panel E).[4] The hazard rate decreases drastically within the first 3 minutes and remains relatively stable afterward. Similar patterns are observed for half-hour shows. Therefore, we categorize viewing durations between 30 seconds and 3 minutes as sampling events (a definition we use throughout this article).[5] Based on this categorization, Figure 1, Panel F, indicates that in 70% of the cases a household decides to watch the first show sampled. Hence, sampling is informative of preferences.

In summary, the data suggest that the frequency of prime-time viewing is high, that households spend considerable time watching in the evening, and that when viewing, they tend to sample shows until finding one of interest. With this characterization of show viewing in mind, we turn to the recording decision.

Show recording. Households can record shows in any given day. Across households and days, 3% of the household-day observations are associated with recording only, 53% are associated with viewing only, and 44% are associated with both viewing and recording. Hence, recording is common.

The TiVo STB can store 40 hours of programming. STB program inventories are near capacity (>90%) on most (82%) household-day pairs in the data, implying that households usually must delete one show before recording another. These recording and deletion decisions are therefore informative about relative show preferences.

Advertising viewing. Advertising viewership is predicated on the series of decisions shown in Figure 2. First, advertising viewing requires one to be watching the show when the advertisement airs (a viewer is exposed). Second, a household must decide whether to watch a show live or recorded. Because most viewing is live, 78% of all advertising

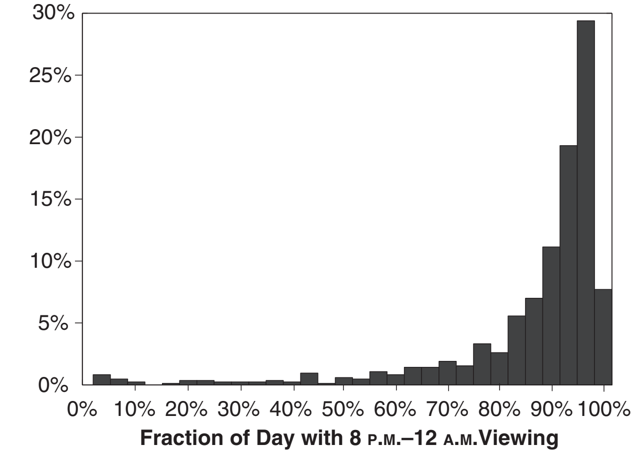
[3]The small spike at 50% occurs because people watching a one-hour show might exit in the middle to watch another show that just started, as most shows start or end around the hour or the half-hour.

[4]The show-switching spikes in Figure 1, Panel E, are coincident with the starting times of shows on other networks. Across all shows, 69.2% start on the hour and 26.9% start on the half-hour, comprising the bulk of starting times. However, some shows commence at 5 minutes, 15 minutes, 20 minutes, 35 minutes, 40 minutes, and 45 minutes after the hour, with respective percentages of .18%, .31%, .14%, 1.16%, .18%, and .37%.

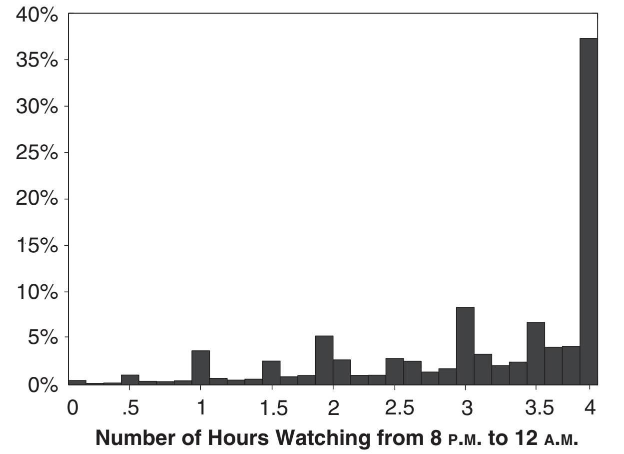
[5]A sensitivity analysis on sampling threshold duration using alternative lower bounds of 10 seconds, 20 seconds, and 60 seconds indicates the number of shows sampled varies less than ±7% relative to the current 30-second threshold.

Figure 1
DESCRIPTIVES ON TV VIEWING

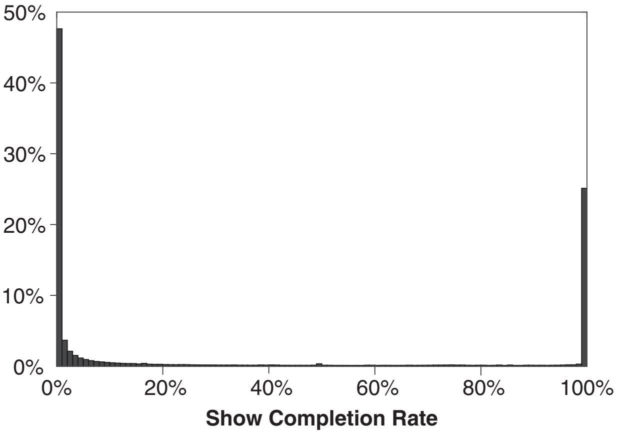
A: Days with 8 P.M.–Midnight Viewing (by Household; n = 834)



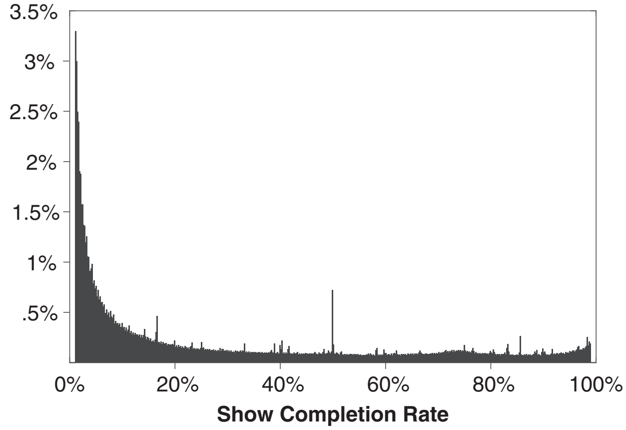
B: Cumulative Daily TV Usage, 8 P.M.–Midnight (by Household-Day; n = 304, 410)



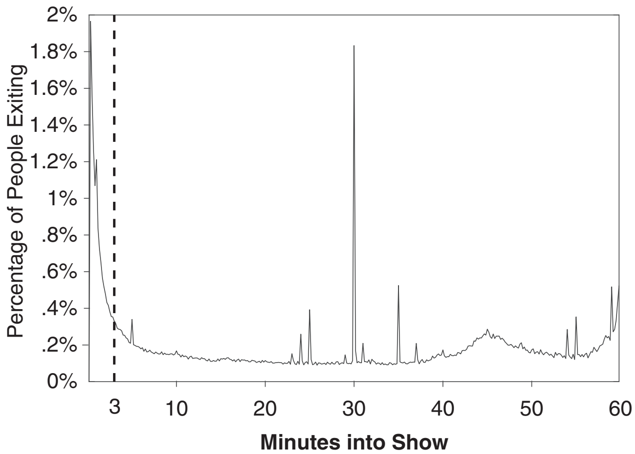
C: Show Completion Rate, All 8 P.M.–Midnight Viewing (by Household-Show; n = 6,617, 351)



D: Show Completion Rate, 8 P.M.–Midnight Viewing with Rate in 1%–99% Range (by Household-Show; n = 6,617, 351)



E: Hazard Rate for One-Hour Shows Watched from Broadcast Start Time (Across Household-Show; n = 611, 571)



F: Number of Shows Sampled Before Watching One (by Household-Show, n = 2,899, 749)

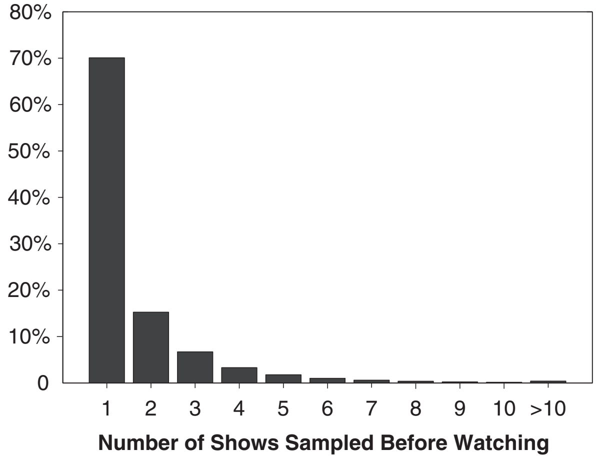
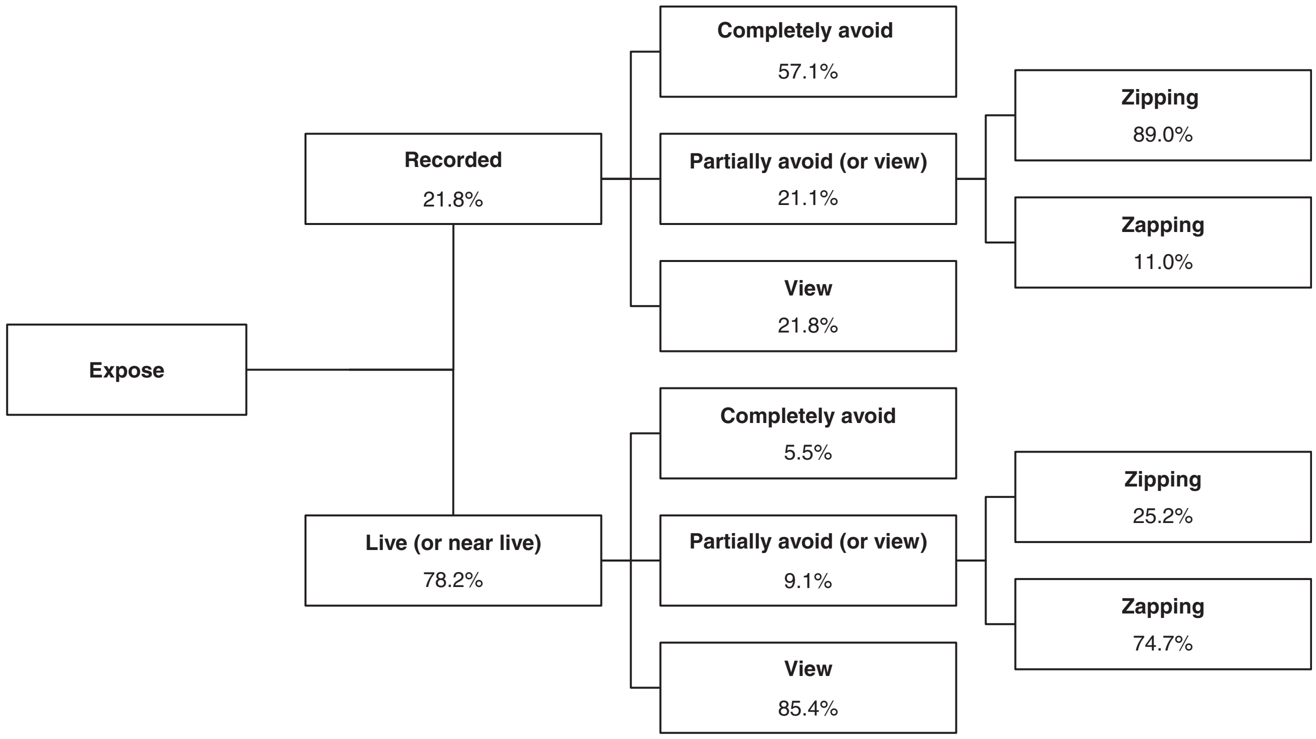


Figure 2
ADVERTISING VIEWING AND AVOIDANCE

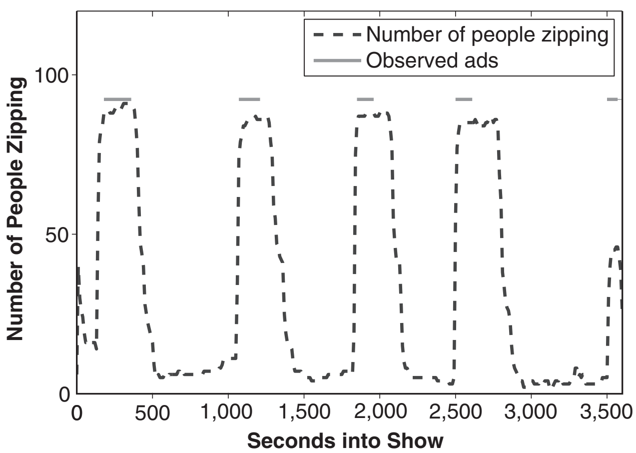


exposures are live. Third, when confronted with an advertisement, a household can decide to avoid it. Complete avoidance occurs when a fast-forward starts before the advertisement and ends after it. Partial avoidance occurs when the household starts or stops fast-forwarding (i.e., zipping) during the advertisement and/or switches channels into or out of the advertisement (i.e., zapping). While zapping occurs in both live and recorded shows, zipping can only occur in recorded or near-live shows (i.e., when the current channel's broadcast is automatically recorded by TiVo). As expected, recorded shows are more subject to avoidance, as evidenced by the nearly 80% avoidance rate (mostly zipping) for recorded shows and a lower than 15% avoidance rate (mostly zapping) for live or near-live shows.

To further exemplify avoidance patterns in recorded shows, we depict in Figure 3 the timing of zips and commercial breaks during a one-hour-long recorded episode of *CSI: Crime Scene Investigation* on December 8, 2005. Figure 3 indicates that zips coincide with commercial breaks.[6] Similar patterns exist in other shows. The figure suggests that advertising avoidance is not uncommon but that when viewers are watching recorded shows, viewership tends to return to the same level after the commercial break.

To ascertain the factors that explain variation in prime-time advertising avoidance from zipping, we conduct a variance decomposition with the following explanatory variables: household, brand, show genre, network, product category, location of the commercial break within the show (i.e., pod) and the slot within the break (i.e., pod position or slot), day of the week, hour, and past avoidance (Table 1). If all the variation in skipping can be apportioned to shows or time, then standard aggregate methods of targeting based on purchasing slots in shows should be effective tools to address

Figure 3
NUMBER OF PEOPLE ZIPPING (BY SECOND) IN ONE EPISODE OF *CSI*



[6]Zipping durations exceed the durations of the national advertising breaks because the TNS advertising schedule data do not include local cable advertisements or advertisements for upcoming cable network programs ("promos" or "tune-ins"), and these are aired at the beginning or the end of the commercial break (Wilbur, Xu, and Kempe 2013).

Table 1
ANALYSIS OF VARIANCE FOR ADVERTISING SKIPPING

Source	*d.f.*	*Type I Sum of Squares*	*Mean Square*	*F-Value*	*Pr > F*	*Percent Variance*
Household	777	113,065.5	145.5	1,160.6	<.0001	20.4%
Brand	1,920	1,586.4	.8	6.6	<.0001	.3%
Genre	14	4,135.9	295.4	2,356.2	<.0001	.7%
Network	55	2,129.1	38.7	308.7	<.0001	.4%
Product category	573	168.6	.3	2.4	<.0001	.0%
Pod	27	3,964.1	146.8	1,171.0	<.0001	.7%
Pod position (slot)	33	415.7	12.6	100.5	<.0001	.1%
Day of week	6	83.8	14.0	111.5	<.0001	.0%
Hour	3	114.0	38.0	303.0	<.0001	.0%
Prior ad avoided	1	65,691.3	65,691.3	523,923.0	<.0001	11.9%

Notes: Data are for recorded shows watched between 8 P.M. and midnight; N = 6,401,894.

avoidance. If, in contrast, there remains substantial household-specific variation, then the efficacy of microtargeting is amplified.

We find the latter to be the case. The factors incorporated in the model account for 34.5% of the overall variance in advertising viewing and skipping. Most importantly, household-fixed effects account for 59.1% of the explained variance. Moreover, demographic variables alone are not sufficient in explaining the variation. If household-fixed effects are replaced by a set of demographic variables, the total explained variance drops from 34.5% to 27.8%, and observed demographic variables account for only 11.4% of the explained variance. All in all, these results suggest demographic-based advertising buys can be augmented with household-specific advertising avoidance information.

Advertising costs. Table 2 provides summary statistics of per-exposure price for 15-second advertising slots for the shows discussed in the "Nielsen TV Viewing Data" section. The median per-exposure price is about 1 cent for ABC, CBS, and NBC and is slightly above 1 cent for FOX. Moderate price variation also exists within each network.

To ascertain whether advertising prices relate to viewership, we regress the advertisements' prices (in $1,000 units) on ratings and network dummies. Results indicate a positive and significant relationship between price and rating (7.8; SE = .9), and the network coefficient is the highest for FOX (32.9; SE = 3.6), followed by CBS (25.3; SE = 4.7), NBC (16.8; SE = 4.0), and ABC (12.7; SE = 3.6). Presumably, the differences across networks relate to the demographics of the shows viewed (e.g., Goettler 2012).

TV VIEWING MODEL

Summary of Model Components

This summary overviews the various model components used in our targeting analysis. First, we consider the television and advertising viewing model. This modeling component enables us to predict advertising views. Second, we consider the advertising response model in the "Advertising Response" section. This modeling component enables us to link advertising views to sales. Combined, the first two model components enable us to ascertain how changes in advertising buys affect sales. Third, using the link between advertising response and costs, we conduct counterfactual targeting analyses to improve the efficiency of media buys. This modeling component is discussed in the "Targeting Approaches" section. Figure 4 overviews these model components and enumerates data requirements.

Overview of the Viewing Model

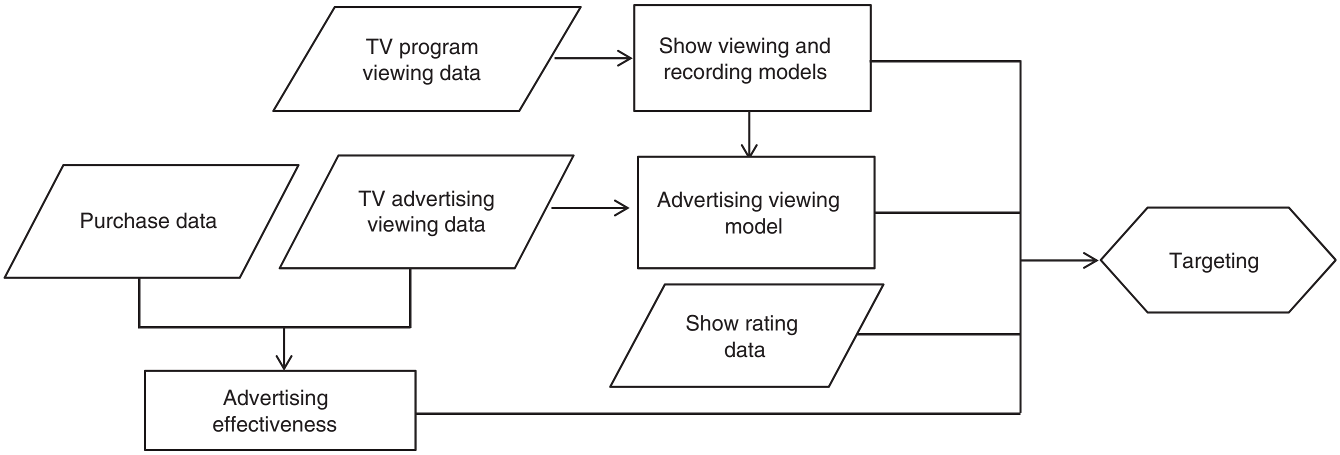
The TV viewing model comprises three components motivated by the preceding empirical discussion: TV show sampling and watching (the "TV Show Viewing" section), TV show recording (the "TV Show Recording" section), and advertising viewing (the "TV Advertising Viewing" section). All three components are predicated upon the theoretical concept of flow utility, that is, the moment-by-moment consumption benefit a viewer derives from watching a TV show, watching an advertisement, or engaging in a non-TV activity (the outside good). In this flow utility framework, viewers derive utility from viewing a show, but they experience ex ante uncertainty about the flow utility of TV shows they are considering. For instance, prior to tuning into an episode of a show, viewers might be uncertain about the story line, the role of a favorite actor, whether the show is rerun, and so forth. This uncertainty induces viewers to sample shows prior to viewing them, watching a candidate show for a short time (sampling) to learn about its quality, eventually settling in on a show when the expected benefit of sampling another show is lower than the effort involved in sampling it.

At random points after choosing a show to watch, the flow utility of the show changes (e.g., with a new scene or show segment), and the viewer evaluates whether to continue watching the show or to sample anew. Likewise, when advertisements are encountered, consumption utility suddenly changes, again leading to the potential for viewers to tune away or fast-forward. Eventually, viewing sessions conclude,

Table 2
SUMMARY STATISTICS OF PER-EXPOSURE ADVERTISING PRICE BASED ON NIELSEN RATINGS (JULY 2006)

Network	*Number of Observations*	*M ($)*	*Mdn ($)*	*SD ($)*
ABC	94	.011	.010	.005
CBS	102	.012	.011	.004
NBC	93	.011	.010	.006
FOX	87	.015	.014	.010

Figure 4
MODEL OVERVIEW



at the end of the evening or when the consumption utility falls short of the outside good.

In the next section, we introduce the three flow utilities: show, advertisement, and outside good. Based on these flow utilities, we then describe the viewer choice process (show sampling, show viewing, show recording, and advertising viewing) in the sections that follow.

Flow Utilities

We model flow utilities of TV shows, TV advertisements, and the outside good. All utilities are viewer specific.

Show utility. Each show can be represented by a unique combination of t (time) and n (network). The flow utility that viewer i derives from show tn is defined as

$$u^S_{itn} = X^S_{itn}\beta^S_i + \nu^S_{itn} + \varepsilon^S_{itn} \equiv \bar{u}^S_{itn} + \nu^S_{itn} + \varepsilon^S_{itn}, \tag{1}$$

where X^S_{itn} is a vector that captures show characteristics and viewer i's past viewing behavior, including genre, network, show length, percentage of show aired when sampling begins (i.e., viewing offset), number of previous episodes of the program that the viewer has sampled in the preceding week, and whether the viewer was watching network n before the current choice occasion. The last two terms capture state dependence in viewing behavior: ν^S_{itn} represents a viewer-show specific error term observed by the viewer but not by the researcher, and ε^S_{itn} represents viewer uncertainty pertaining to program and episode quality prior to sampling that is revealed to the viewer only after sampling the show. We assume ν^S_{itn} is i.i.d. standard Type I extreme value distributed and ε^S_{itn} is i.i.d. Type I extreme value distributed with mean 0.[7]

[7]To accommodate the potential that viewer experience might reduce this uncertainty prior to sampling, we interact past viewing behavior with the remaining model covariates. Web Appendix A shows that this approach is equivalent to allowing for heterogeneous uncertainty based on differential experience with shows. We thank an anonymous reviewer for motivating this analysis.

Advertisement utility. Each advertisement insertion can also be represented by a unique combination of t (time) and n (network). The flow utility that viewer i obtains from advertisement tn is a function of the characteristics of the advertisement X^A_{itn} and is given by

$$u^A_{itn} = X^A_{itn}\beta^A_i + \varepsilon^A_{itn} \equiv \bar{u}^A_{itn} + \varepsilon^A_{itn}, \tag{2}$$

where X^A_{itn} includes pod, pod position (slot), genre of the associated show, product category, and whether the preceding advertisement is avoided. The term ε^A_{itn} is an idiosyncratic error term affecting the inherent valuation of advertisement tn (conditional on exposure) and is observed by the viewer but not by the researcher. We assume ε^A_{itn} to be i.i.d. standard Type I extreme value distributed.

Outside good utility. When allocating time, the viewer contrasts the utility from viewing TV to the best available alternative (i.e., the outside good). If the utility from viewing TV exceeds that of the outside good, the viewer will watch TV. As such, the flow utility of the outside good is tantamount to the "opportunity cost" of time, posited to vary by viewer (i) and time (t) and denoted as:

$$u^O_{it} = X^O_{it}\beta^O_i + \nu^O_{it} \equiv \bar{u}^O_{it} + \nu^O_{it}, \tag{3}$$

where X^O_{it} is a vector of observable day characteristics specific to viewer i and time t. These characteristics include weekday fixed effects and indicators for previous-day TV viewing and previous-weekday TV viewing. Monthly fixed effects are added to control for seasonality. The term ν^O_{it} is an idiosyncratic error term affecting the utility from the outside good at time t and is observed by the viewer but not by the researcher. We assume ν^O_{it} to be i.i.d. standard Type I extreme value distributed and i.i.d. with $\nu^S_{itn}(\forall n)$.

TV Show Viewing

Owing to ε^S_{itn} in Equation 1, viewers face ex ante uncertainty in the utility of viewing that can only be resolved by sampling a show (that is, briefly viewing a show). We assume that a viewer first samples the alternative with the highest ex ante expected viewing utility given the observed show characteristics and ν^S_{itn}, the shock unobserved to the researcher. This ordering of alternatives is simple for the viewers to do (suggesting the process is a parsimonious representation) and is also consistent with the optimal search ordering implied by Weitzman (1979) and Kim, Albuquerque, and Bronnenberg (2010) when uncertainty is large relative to search costs. After the short exposure to the show, the viewer observes the utility shock for that show, ε^S_{itn}, and decides whether to continue watching by comparing the flow utility of that show with the expected highest flow utility to be obtained from other shows available at that time (for which the $\varepsilon^S_{itn'}$ at the other shows tn′ is still unknown to the viewer). If the flow utility of the current show is lower, the viewer samples other shows.

If the flow utility of the current show is higher than the expected best remaining alternative, the viewer selects the show and enters a "flow" state of watching. During this flow state, the flow utility of the show remains constant until an external "shock" randomly arrives that changes the flow utility by perturbing ε^S_{itn}. Such a perturbation might reflect a change in a story on a news show, for example. When this occurs, a new error is drawn in the flow utility model in Equation 1. If the resulting flow utility is lower, the viewer compares this new utility with the expected highest flow utility that could be obtained from switching to another show or the outside good. If these alternatives yield higher utility, the viewer switches. An analogous process holds for advertising. When an advertisement appears within the show, the flow utility changes to that of the advertisement, and the viewer again compares the flow utility with other options. Figure 5 overviews this process and the sections of this article that elaborate on the show sampling and viewing decisions (while the ad viewing process is described in the "TV Advertising Viewing" section). In this figure, ε^S_{itn} is revealed in the "Sample show" step. We elaborate on the sampling and viewing processes next.

Sampling. Viewers sample shows to ascertain whether the flow utility is likely to exceed that of other options. At time t, the set of available shows to be sampled consists of all current live programs and the current menu of recorded programs. The viewer starts by sampling the show with the highest expected flow utility. Prior to the viewer observing ε^S_{itn} ($\forall$n), the expected flow utility of show tn is $\bar{u}^S_{itn} + \nu^S_{itn} + E(\varepsilon^S_{itn})$. Because the ε^S_{itn} ($\forall$n) are i.i.d. distributed with mean zero, $E(\varepsilon^S_{itn})$ is equal across shows, and the sampling decision is therefore incumbent upon $\bar{u}^S_{itn} + \nu^S_{itn}$. Show tn is sampled if $\bar{u}^S_{itn} + \nu^S_{itn} > \bar{u}^S_{itn'} + \nu^S_{itn'}, \forall n'$ and $\bar{u}^S_{itn} + \nu^S_{itn} > \bar{u}^O_{it} + \nu^O_{it}$. Under the assumption that ν^S_{itn} ($\forall$n) and ν^O_{it} in Equations 1 and 3 are i.i.d. standard Type I extreme value distributed, this probability is given by

$$\Pr\left(y^S_{itn} = 1\right) = \frac{\exp\left(\bar{u}^S_{itn}\right)}{\exp\left(\bar{u}^O_{it}\right) + \sum_{n'}\exp\left(\bar{u}^S_{itn'}\right)}, \tag{4}$$

where y^S_{itn} is an indicator that show tn is sampled by viewer i.

After sampling show tn, viewer i observes ε^S_{itn} and therefore u^S_{itn}. The viewer then compares u^S_{itn} with the expected highest flow utility (i.e., the inclusive value) to be obtained from the remaining nonsampled shows available at that time, as well as the outside good:

$$\begin{aligned}
&IV_{it\mathcal{N}_{it}}\left|\left\{\nu^S_{itn}\right\}_n, \nu^O_{it}\right.\\
&\quad = E_{\left\{\varepsilon^S_{itn'}\right\}_{n'\in\mathcal{N}_{it}}} \max_{n'\in\mathcal{N}_{it}}\left\{\bar{u}^S_{itn'} + \nu^S_{itn'} + \varepsilon^S_{itn'}, \bar{u}^O_{it} + \nu^O_{it}\right\}\\
&\quad = \max\left\{\bar{u}^O_{it} + \nu^O_{it}, \ln\left(\sum_{n'\in\mathcal{N}_{it}}\exp\left(\bar{u}^S_{itn'} + \nu^S_{itn'}\right)\right)\right\},
\end{aligned}$$

where $\mathcal{N}_{it}$ denotes the set of networks at time t that has yet to be sampled.

If $u_{itn} \geq IV_{it\mathcal{N}_{it}}$, viewer i watches the show. The probability of this event is given by

$$\begin{aligned}
&\Pr\left(y^W_{itn} = 1\left|y^S_{itn} = 1, \left\{\nu^S_{itn}\right\}_n, \nu^O_{it}\right.\right)\\
&\quad = \Pr\left(u^S_{itn} \geq IV_{it\mathcal{N}_{it}}\left|\left\{\nu^S_{itn}\right\}_n, \nu^O_{it}\right.\right)\\
&\quad = 1 - F_{\varepsilon^S_{itn}}\left(\max\left\{\bar{u}^O_{it} + \nu^O_{it},\right.\right.\\
&\qquad \left.\left.\ln\left(\sum_{n'\in\mathcal{N}_{it}}\exp\left(\bar{u}^S_{itn'} + \nu^S_{itn'}\right)\right)\right\} - \bar{u}^S_{itn} - \nu^S_{itn}\right),
\end{aligned} \tag{5}$$

where y^W_{itn} is an indicator that viewer i watches show tn and $F_{\varepsilon^S_{itn}}(\cdot)$ denotes the cumulative distribution function of ε^S_{itn}.

Computation of $\Pr(y^W_{itn} = 1|y^S_{itn} = 1)$ involves integrating out $\{\nu^S_{itn}\}_n$ and ν^O_{it} in $\Pr(y^W_{itn} = 1|y^S_{itn} = 1, \{\nu^S_{itn}\}_n, \nu^O_{it})$. The sampling order implies that $\nu^S_{itn'}$ and ν^O_{it} in Equation 5 are truncated below $\bar{u}^S_{itn} + \nu^S_{itn} - \bar{u}^S_{itn'}$ and below $\bar{u}^S_{itn} + \nu^S_{itn} - \bar{u}^O_{it}$, respectively.

If $u^S_{itn} < IV_{it\mathcal{N}_{it}}$, viewer i samples another show if the inclusive value of remaining shows is higher than the value of the outside good. The choice set is now $\tilde{\mathcal{N}}_{it} = \mathcal{N}_{it}\backslash n$, and the probability of sampling show $t\tilde{n}$ ($\tilde{n} \in \tilde{\mathcal{N}}_{it}$) is

$$\Pr\left(y^S_{it\tilde{n}} = 1\right) = \frac{\exp\left(\bar{u}^S_{it\tilde{n}}\right)}{\exp\left(\bar{u}^O_{it}\right) + \sum_{n'\in\tilde{\mathcal{N}}_{it}}\exp\left(\bar{u}^S_{itn'}\right)}. \tag{6}$$

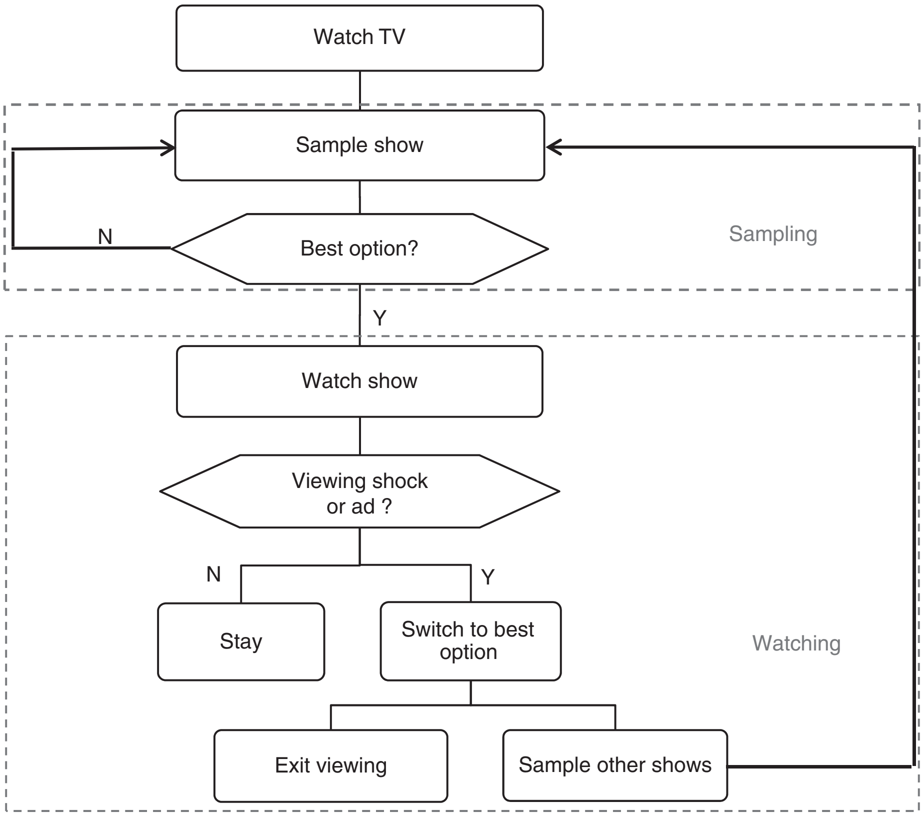
The sampling process repeats until either the viewer identifies a show that is worth watching (in which case the viewer watches the show), or the value of the outside good exceeds the inclusive value of remaining shows (in which case the viewer ends the viewing session).

Watching. Upon selecting show tn to view, viewer i obtains viewing flow utility u^S_{itn} until the show ends or the viewer exits viewing, whichever comes first. The decision to stop watching is driven by arrival of external shocks that change the flow utility (Arcidiacono et al. 2015; Nevskaya and Albuquerque 2012). If the external shock is sufficiently negative, the viewer terminates the show. This characterization of the show exiting decision is in essence similar to the "first hitting time" models (Lee and Whitmore 2006).

Specifically, at some time $t' > t$, viewer i encounters an external shock $\varepsilon^S_{it'n}$ (e.g., change in plot, actor, or scene), which replaces ε^S_{itn} and changes the flow utility of show tn from $\bar{u}^S_{itn} + \nu^S_{itn} + \varepsilon^S_{itn}$ to $\bar{u}^S_{itn} + \nu^S_{itn} + \varepsilon^S_{it'n}$. If this new flow utility falls below the inclusive value of remaining alternatives, the viewer will exit the show.

To characterize the duration until the arrival of a new external shock, we assume a homogeneous Poisson process with rate λ_{itn} for viewer i and show tn; λ_{itn} is parameterized as a function of genre:

Figure 5
DECISION MAKING IN ONE VIEWING SESSION



$$\lambda_{itn} = \exp(g_{tn}\rho_i), \tag{7}$$

where g_{tn} is a row vector on genre, the jth element being an indicator variable of whether show tn is of the jth genre.

Under this Poisson assumption, the probability of viewer i exiting show tn at time t′, $q_{itnt'}$, is given by the probability that the flow utility upon receiving a new viewing shock falls below the alternative options:

$$\begin{aligned} q_{itnt'} \Big| \{\nu^S_{itn}\}_n, \nu^O_{it} &= \Pr\Bigg(\bar{u}^S_{itn} + \nu^S_{itn} + \varepsilon^S_{it'n} < \max \\ &\quad \Bigg\{\bar{u}^O_{it} + \nu^O_{it}, \ln\Bigg(\sum_{n' \in \mathcal{N}_{it'}} \exp\left(\bar{u}^S_{it'n'} + \nu^S_{it'n'}\right)\Bigg)\Bigg\}\Bigg) \\ &= F_{\varepsilon^S_{it'n}}\Bigg(\max\Bigg\{\bar{u}^O_{it} + \nu^O_{it}, \\ &\quad \ln\Bigg(\sum_{n' \in \mathcal{N}_{it'}} \exp\left(\bar{u}^S_{it'n'} + \nu^S_{it'n'}\right)\Bigg)\Bigg\} - \bar{u}^S_{itn} - \nu^S_{itn}\Bigg). \end{aligned} \tag{8}$$

Computation of $q_{itnt'}$ involves integrating out $\{\nu^S_{itn}\}_n$ and ν^O_{it} in $q_{itnt'} \big| \{\nu^S_{itn}\}_n, \nu^O_{it}$; $q_{itnt'}$ is not necessarily fixed through the duration of show tn, and it can change when available shows on alternative networks ($\mathcal{N}_{it'}$) change. For instance, when a new show t′n′ starts on network n′, $\bar{u}^S_{it'n'} + \nu^S_{it'n'}$ changes, and $q_{itnt'}$ will change accordingly. Thus, $q_{itnt'}$ is piecewise constant and changes whenever shows on alternative networks ($\mathcal{N}_{it'}$) change. Note that between shocks, $q_{itnt'}$ remains fixed.

Using an approach similar to those of Arcidiacono et al. (2015) and Nevskaya and Albuquerque (2012), Web Appendix B shows that for $q_{itnt'}$ that is piecewise constant with segment 1, ..., M, the cumulative distribution function of the viewing length l^*_{itn} is

$$F_{l^*_{itn}}(\bar{t}) \equiv \Pr\{l^*_{itn} \le \bar{t}\} = 1 - e^{-\lambda_{itn}\sum_{m=1}^{M} l^m_{itn} q^m_{itn}}, \tag{9}$$

where q^m_{itn} is the exiting probability in segment m and l^m_{itn} is the length of segment m up to time $\bar{t}$, $\sum_{m=1}^{M} l^m_{itn} = \bar{t}$.

The probability density function of l^*_{itn} is therefore

$$f_{l^*_{itn}}(\bar{t}) = \lambda_{itn} q^{\bar{m}}_{itn}\, e^{-\lambda_{itn}\sum_{m=1}^{M} l^m_{itn} q^m_{itn}}, \tag{10}$$

where $\bar{m}$ is the segment that $\bar{t}$ falls into.

Because it is not possible to watch a show past its end,

$$l^W_{itn} = \min\left(l^*_{itn}, L_{itn}\right), \tag{11}$$

where l^W_{itn} is the time viewer i spends watching show tn and L_{itn} is the remaining length of show tn when the show is sampled.

TV Show Recording

The TiVo STB used by the panelists records one show at a time. Therefore, a newly recorded show tn is assumed to have

(1) higher expected flow utility than the show that is replaced ($u^S_{itn} > u^S_{itd}$) and (2) higher expected flow utility than all shows that air at time t but are not recorded ($u^S_{itn} > u^S_{itn'}, \forall n' \neq n$). Because the show deleted is nearly always automatically selected by the STB, we refrain from using the deletion choice decision.

Based on the flow utility specified in Equation 1, condition 2 implies the probability that viewer i records show tn is given by

$$\Pr\left(y^R_{itn} = 1\right) = \frac{\exp\left(\bar{u}^S_{itn}\right)}{\sum_{n'} \exp\left(\bar{u}^S_{itn'}\right)}, \tag{12}$$

where y^R_{itn} is an indicator that show tn is recorded by viewer i.

TV Advertising Viewing

TV advertising avoidance differs between live and recorded viewing. In live viewing, advertising avoidance involves switching away from the advertisement (zapping), so the viewer's alternative set includes other shows. In contrast, almost all advertising avoidance when views are recorded involves forwarding (zipping), so in this case, the viewer's alternative set includes the opportunity cost of time. We formalize these points next.

Zapping. In the viewing model introduced above, the viewer can choose to avoid advertisements in a live show by channel switching (zapping). Similar to show viewing in which viewers can make channel switching decisions during program content upon receiving an external shock, viewers make zapping decisions during each advertisement. Hence, this component of the viewing model is analogous to the TV show watching model, except we observe the specific point at which the utility changes. The zapping decision depends on the relative attractiveness of the advertisement versus shows on alternative networks and the outside good, and the cost of zapping. The probability of zapping the live (L) advertisement tn can be written as

$$\begin{aligned} &\Pr\left(y^{AL}_{itn} = 0 \middle| \left\{\nu^S_{itn}\right\}_n, \nu^O_{it}\right) \\ &= \Pr\left(\bar{u}^A_{itn} + \varepsilon^A_{itn} < \max \left\{\bar{u}^O_{it} + \nu^O_{it}, \ln\left(\sum_{n' \in \mathcal{N}_{it}} \exp\left(\bar{u}^S_{itn'} + \nu^S_{itn'}\right)\right)\right\} - c_i\right) \\ &= F_{\varepsilon^A_{itn}}\left(\max\left\{\bar{u}^O_{it} + \nu^O_{it}, \ln\left(\sum_{n' \in \mathcal{N}_{it}} \exp\left(\bar{u}^S_{itn'} + \nu^S_{itn'}\right)\right)\right\} - c_i - \bar{u}^A_{itn}\right), \end{aligned} \tag{13}$$

where y^{AL}_{itn} is an indicator that the live advertisement tn is viewed (i.e., not zapped) by viewer i and c_i is the zapping cost faced by viewer i.[8] Computation of $\Pr(y^{AL}_{itn} = 0)$ involves integrating out $\{\nu^S_{itn}\}_n$ and ν^O_{it} in $\Pr(y^{AL}_{itn} = 0 | \{\nu^S_{itn}\}_n, \nu^O_{it})$.

Zipping. Because zipping reduces viewing time, the zipping decision is reached by comparing the flow utility of the advertisement with that of the outside good. If the advertisement provides higher flow utility than the outside good, the viewer watches it. The probability that viewer i zips a recorded (R) advertisement tn is

$$\Pr\left(y^{AR}_{itn} = 0\right) = \frac{\exp\left(\bar{u}^O_{it}\right)}{\exp\left(\bar{u}^O_{it}\right) + \exp\left(\bar{u}^A_{itn}\right)}, \tag{14}$$

where y^{AR}_{itn} is an indicator that the recorded advertisement tn is viewed (i.e., not zipped) by viewer i.[9]

The observed show sampling and watching decisions, show recording decisions, and advertising viewing decisions enable us to recover flow utilities of TV shows, TV advertisements, and the outside good. We discuss estimation in the next section.

VIEWING MODEL ESTIMATION AND RESULTS

Estimation

The viewing model is estimated by simulated maximum likelihood. The likelihood, derived in Web Appendix C, is the product of likelihoods associated with sampling, watching conditional on sampling, viewing length, recording, zapping, and zipping. We discuss model identification in Web Appendix C.

All parameters are viewer specific, as indicated by subscript i. We perform estimation viewer by viewer, facilitated by the availability of panel data of relatively long cross-section and duration for each viewer.[10] Estimation data cover July 2005 to June 2006, while data from July 2006 is reserved for the policy experiments.[11]

Because virtually every viewer watches only a handful of available networks, we construct viewer-specific consideration sets on the basis of viewing history. For each viewer, the consideration set of networks consists of the smallest number of networks that collectively account for at least 90% of prime-time viewing time. On each viewing occasion, the choice set comprises the following two types of shows: (1) live shows that are available on networks within the consideration set and (2) shows stored on a STB that are recorded either manually or through a season pass where episodes of a given show is automatically recorded.

To limit the size of parameters governing show and advertising flow utilities (β^S_i and β^A_i), we only estimate flow utility parameters associated with the six most popular genres (drama, comedy, reality TV, talk shows, news, and sports, together accounting for 68% of viewing time) and the six most popular networks (ABC, CBS, NBC, FOX, USA, and Comedy Central, together accounting for 60% of viewing time). There are 574 product categories in the advertising data. In an initial effort to capture the effects of product category on advertising preference, we classify the 574 product categories into four

[8]We consider two zapping costs: zapping during the show and zapping during the advertisement. Zapping costs can reflect the cognitive cost associated with channel switching. For shows, this zapping cost cannot be separately identified from the arrival rate of external shocks. A viewer who switches channels less often during a program can either have a low shock arrival rate or a large channel-switching cost. However, for advertisements, the difference between zipping rates and zapping rates is informative about the relative zipping and zapping costs, as we discuss in the next section.

[9]Theoretically, there can be a cost associated with zipping. However, it cannot be separately identified from the utility of the outside good, so we normalize the zipping cost to zero. Thus, the zapping cost essentially measures the relative cost of zapping versus zipping, and it is identified from the difference in zipping and zapping probabilities.

[10]On average, there are 1,313 sampling occasions and 1,993 advertising viewing occasions per viewer.

[11]We performed several model validity checks using the holdout sample of July 2006. Detailed results are available in Web Appendix D.

Table 3

FLOW UTILITY PARAMETER ESTIMATES FOR SHOWS (β^S) AND ADVERTISEMENTS (β^A)

Variable	*Show*	*Advertisement*
Show length	–.20 (1.2%, 65.7%)	
Episodes sampled in preceding week	.83 (49.5%, 15.9%)	
Lag network	.63 (67.6%, 1.9%)	
Live	9.79 (79.9%, .0%)	
Viewing offset	–4.96 (.0%, 82.0%)	
Prior ad viewed		5.23 (66.6%, .6%)
First ad break		–.11 (6.4%, 16.3%)
Last ad break		.05 (11.3%, 10.8%)
First slot in a break		1.51 (55.3%, 1.0%)
Last slot in a break		–.11 (3.8%, 7.3%)
Genre		
Drama	–.21 (12.9%, 34.2%)	–.77 (9.0%, 25.9%)
Comedy	–.47 (8.2%, 48.2%)	–.10 (8.8%, 15.2%)
Reality TV	–.74 (6.4%, 48.4%)	–.78 (8.6%, 24.7%)
Talk shows	.26 (34.7%, 15.9%)	.30 (7.6%, 10.2%)
News	.23 (32.3%, 12.8%)	–.10 (7.8%, 11.4%)
Sports	–.25 (13.7%, 24.6%)	–.15 (7.3%, 15.2%)
Network		
ABC	.81 (45.6%, 11.5%)	–.03 (12.5%, 16.7%)
CBS	.33 (33.6%, 20.0%)	–.14 (10.9%, 15.3%)
NBC	.50 (40.3%, 19.3%)	–.18 (11.2%, 17.5%)
FOX	.24 (24.9%, 15.2%)	–.14 (10.0%, 12.9%)
USA	–.09 (9.0%, 10.8%)	–.22 (6.0%, 6.8%)
Comedy Central	–.12 (8.4%, 8.2%)	.78 (4.4%, 2.0%)
Product Category		
CPG		–.02 (4.0%, 4.2%)
Service		.01 (4.8%, 3.6%)
Drug		–.09 (2.0%, 5.3%)

Notes: Data shown are median parameter estimates, followed by the percentages of viewers with significant positive and negative estimates (5% significance).

general categories: CPG, services, pharmaceuticals, and other goods.

Results

Table 3 reports the estimation results of the viewing model. The second column reports the median (across viewers) of parameter estimates in flow utilities of shows as well as the percentages of viewers with significant positive and negative estimates (5% level). For most viewers, shorter shows are associated with higher utility: 65.7% of people have a significant negative coefficient on "length." As expected, the average viewer prefers familiar shows, indicating a positive effect of episodes sampled in the preceding week ("experience") on show utility. State dependence is evidenced in network viewing, as indicated by the generally positive coefficients on "lag network." The coefficients for "live" indicate the average viewer prefers live shows over recorded shows. The negative coefficient for "viewing offset" (which measures how far into a show a viewer starts watching) implies that shows with less time elapsed from the start of the show are preferred. In other words, viewers prefer to watch shows in which they have missed less content.

The third column of Table 3 reports the median (across viewers) of parameter estimates in flow utilities of advertisements, as well as the percentages of viewers with significant positive and negative estimates (5% level). The coefficients for "first slot in a break" indicate that the first advertisement in a commercial break is less likely to be avoided, presumably because it takes viewers some time to initiate an avoidance action. Most viewers have a positive coefficient for preceding commercial viewed, indicating state dependence in advertising viewing and implying that viewers forward blocks of advertisements successively.

There also exists extensive heterogeneity in zapping cost across viewers, meaning some viewers do not avoid live advertisements (and presumably are better targets than those who do). The mean, median, and standard deviation of the zapping cost (c) are 2.7, .9, and 4.7, respectively, and the 2.5% and 97.5% quantiles are.01 and 2.9, respectively.

Finally, the average estimated number of shocks per hour (λ) is .49, .66, 1.41, .63, 1.30, and 1.22 for drama, comedy, reality TV, talk shows, news, and sports, respectively. The average viewer is more likely to switch channels during news and less likely during drama shows, perhaps due to differences in program continuity; for example, news broadcasts are frequently punctuated by new stories. This finding is consistent with Shachar and Emerson's (2000) finding that viewing persistence is higher for dramas and lower for news and sports.

Overall, there exists considerable heterogeneity across viewers in viewing preferences for genre, network, and advertising. Heterogeneity in viewing preferences, together with heterogeneity in advertising response, suggests the potential gains available from advertising targeting.

ADVERTISING TARGETING

This section describes advertising targeting policy experiments. We first link advertising viewing with sales response. Predicated on viewing behavior and advertising response, we discuss various targeting strategies.

Advertising Response

To measure advertising response, we select seven product categories that evidence high variation in advertising and sales: children's yogurt, children's cereal, regular cola, diet cola, sports drink, toothpaste, and bathroom tissue. These categories are regularly purchased and frequently advertised, with high cross-sectional and temporal variation in both purchase and advertising. Within these categories, we consider 22 leading brands that have nonnegligible unit market share and that advertised during the sample period. For more detail on these purchase data, refer to Web Appendix E.

Within any given category, the utility household i obtains from purchasing brand j (j = 1, ..., J) of category c in shopping trip m is given by

$$U^{P}_{ijm} = Z_{ijm}\theta_{ic} + h_{ic}(A_{ijm}) + \varepsilon^{P}_{ijm}, \tag{15}$$

where Z_{ijm} is a vector that includes a brand fixed effect, the price of brand j at shopping trip m, an indicator of whether brand j is on promotion (either display or feature) at shopping trip m, and an indicator of whether household i purchased brand j in the previous category purchase. The function $h_{ic}(A_{ijm})$ captures the advertising effect.[12] Finally, ε^{P}_{ijm} is an idiosyncratic error term affecting the inherent valuation of brand j at shopping trip m and is observed by the household but not by the researcher.

The function $h_{ic}(A_{ijm})$ is assumed to be linear in advertising views (though advertising effects are not linear owing to the logit-based demand system we use):

$$h_{ic}(A_{ijm}) = \sum_{a=1}^{A} N^{A}_{ijam}\gamma_{ijac} + \left(N^{A}_{ijm} - \sum_{a=1}^{A} N^{A}_{ijam}\right)\gamma_{ij0c}, \tag{16}$$

where $a \in \{1, ..., A\}$ indexes advertising creatives (as defined in the "TNS Advertising Schedule Data" section) and N^{A}_{ijam} is household i's number of views of advertising creative a since the previous category purchase or in the past seven days, whichever is shorter.[13] We combine creatives that have fewer views because they are individually likely to have a negligible effect on demand. In particular, if N^{A}_{ijm} is household i's total number of advertising views across brand j's creatives during this window, then $N^{A}_{ijm} - \sum_{a=1}^{A} N^{A}_{ijam}$ represents the total number of views on all other creatives with fewer views. Equation 16 allows us to measure creative-specific effects for major creatives (e.g., Malaviya, Kisielius, and Sternthal 1996; Tellis et al. 2005) while pooling the effect of the smaller creatives.

The utility associated with the outside good (i.e., no purchase) is given by

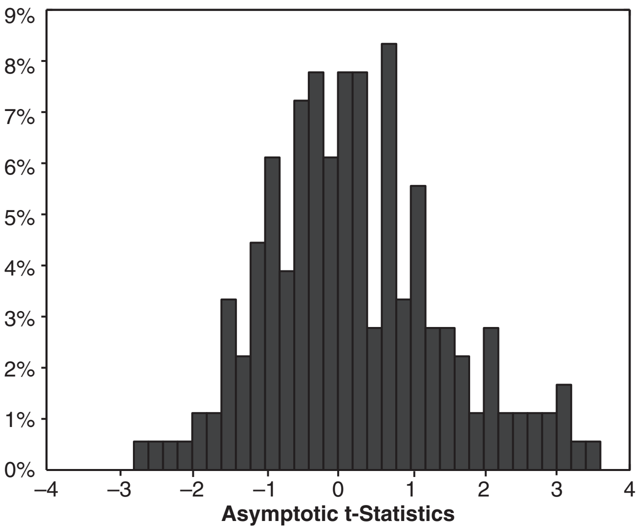
$$U^{P}_{i0m} = \varepsilon^{P}_{i0m}. \tag{17}$$

[12]Firms may target households that are more responsive to advertising. To address this potential endogeneity concern, we compute, for each household and each brand, (1) the brand's share in the household's category purchase and (2) the brand's share in the household's category advertisement exposure. We do not find a correlation between the two variables, implying the advertising variables are unlikely to be endogenous. This lack of correlation might be a consequence of the current targeting practice, and it implies the potential for improvement. In addition, the between-household variance of advertising exposure is lower than the within-household variance. For the 22 brands in the sample, the mean and median portions of the total variance between households are, respectively, 20.1% and 15.9%, with the lowest being 4.0% and the highest being 47.9%.

[13]We use a seven-day window as this typically represents two shopping trips, enhancing the likelihood one can attribute a given view to a purchase decision on a given shopping trip, akin to last touch attribution.

Figure 6

DISTRIBUTION OF THE ASYMPTOTIC T-STATISTICS OF ESTIMATED ADVERTISING COEFFICIENTS



Assuming the idiosyncratic error terms to be i.i.d. standard Type I extreme value distributed, the probability that household i chooses brand j in shopping trip m is

$$\Pr\left(y^{P}_{ijm} = 1\right) = \frac{\exp\left[Z_{ijm}\theta_{ic} + h_{ic}(A_{ijm})\right]}{1 + \sum_{j'=1}^{J}\exp\left[Z_{ij'm}\theta_{ic} + h_{ic}(A_{ij'm})\right]}. \tag{18}$$

Using a latent class model (Kamakura and Russell 1989) to capture household heterogeneity, the probability that household i in segment k (k = 1, ..., K) chooses brand j in shopping trip m is:

$$\Pr\left(y^{P}_{ijm} = 1 \middle| i \in k\right) = \frac{\exp\left[Z_{ijm}\theta_{kc} + h_{kc}(A_{ijm})\right]}{1 + \sum_{j'=1}^{J}\exp\left[Z_{ij'm}\theta_{kc} + h_{kc}(A_{ij'm})\right]}, \tag{19}$$

where $h_{kc}(A_{ijm}) \equiv \sum_{a=1}^{A} N^{A}_{ijam}\gamma_{kjac} + (N^{A}_{ijm} - \sum_{a=1}^{A} N^{A}_{ijam})\gamma_{kj0c}$.

The model is separately estimated for each product category, resulting in a unique set of advertising coefficients across product categories, consumer segments, and advertising creatives. To assess whether advertising affects sales, Figure 6 plots the histogram of the asymptotic t-statistics of the estimated advertising coefficients.[14] As in prior research, advertising effects are statistically small, but when they are distinguishable from zero, they are mostly positive. Were advertising effect zero, the distribution of the asymptotic t-statistics would follow a standard normal distribution (per the central limit theorem). A one-sample Kolmogorov–Smirnov test rejects the hypothesis that the asymptotic t-statistics follow a standard normal distribution ($p < .07$), indicating that advertising effects overall are statistically greater than zero.[15] To assess the

[14]Detailed estimation results are available from the authors.

[15]At the 5% significance level, 19 out of 129 (15%) consumer segment–creative pairs are significantly positive , and 4 (3%) are significantly negative. Thus, negative effects are slightly less than one might expect by chance, whereas positive effects are substantially greater.

Table 4
PERCENTAGES OF SIGNIFICANT ADVERTISING EFFECTS, BY STUDY

Study	p = *.05*	p = *.2*	p = *.4*
Eastlack and Rao (1989)	24%		
Lodish et al. (1995)			55% for new products; 36% for established brands
Hu, Lodish, and Krieger (2007)		39% before 1995; 45% after 1995	
This study	15%	21%	36%

external validity of our findings, we compare the percentage of estimated advertising effects that are significant at different *p*-values with the percentages reported by existing review papers on advertising effects. The results are reported in Table 4 and are generally in line with previous findings.

To assess the magnitude of advertising effects that are significant (at the 10% level), we simulate the effect of removing the associated advertising creative on sales. Following Sethuraman, Tellis, and Briesch (2011) who find, on average, the long-term advertising elasticity is twice as high as the short-term elasticity, we measure both short- and long-term effect. The short-term effect is obtained by simulating contemporaneous changes in market share, holding all else constant. The long-term effect takes into account not only the immediate effect of eliminating the advertising creative but also the carryover effect of altered choice through the purchase event feedback measure (whether the consumer purchased the brand in the previous purchase).[16]

Figure 7 depicts the results in terms of percentage changes in market share for those creatives with statistically significant ad effects. The length of the black segment of each bar represents the short-term effect, and the total length of the bar represents the long-term effect. As shown in the figure, the change is negative for the majority (18 out of 22) of consumer segment–creative pairs. Among such pairs, the combined short-term and long-term changes in own market share due to elimination of creatives vary from −5.5% to −.4%, suggesting the long-term effects of an advertising execution are not insubstantial.

Overall, we conclude that the significant advertising effects are generally positive, but small, as in the previous literature (e.g., Sethuraman, Tellis, and Briesch 2011).[17]

Targeting Approaches

Currently, national TV networks typically sell advertising inventory in advance by show. In the up-front market, advertisers purchase advertising across a show or set of shows for the entire season. Procurement involves a negotiated cost per thousand viewers (CPM), with performance targets for specific periods and programs. While the TV network lists available commercial space by show and air date, the exact location of the advertisement's placement within the show is determined at a later stage. Advertising prices vary with the number of show viewers as well as the demographic mix.

Digital distribution offers two further advances upon the up-front model. First, this technology allows advertisers to buy users instead of shows. As the viewing and advertising avoidance models outlined in the section entitled "TV Viewing Model" enable advertisers to forecast which shows will be viewed by their target audience, this suggests the potential of our approach to enhance the efficiency of advertising buys in up-front markets. Second, advertisements can be inserted in real time, analogous to current practices in Internet advertising. In this case, the advertiser observes all information at the time immediately preceding the available advertising slot. This information set includes the show watched just before the ad airs (and at the time of the commercial break). Thus, in the real-time case, the advertiser knows the TV is on and the consumer is watching a show when the slot appears, whereas in the advance-buy case, these two decisions are known only as far as the model's ability to forecast them.[18]

Several metrics can be employed when targeting. At the simplest level, one can either maximize target views for a given budget or minimize a budget for a given view. The advantage of these metrics is that they do not require a model of advertising response or sales data to implement. In our analysis, we focus on minimizing a budget for a given view. At a more complex level, one can consider the role of profits or revenues. At the cost of invoking an advertising response model and

Table 5
TARGETING SCENARIOS AND ASSOCIATED INFORMATION REQUIREMENTS

	Cost-Based	*Profit-Based*
Real-time buy	• Ad viewing model	• Ad viewing model • Sales response model
Advance buy	• Ad viewing model • Show viewing model	• Ad viewing model • Sales response model • Show viewing model

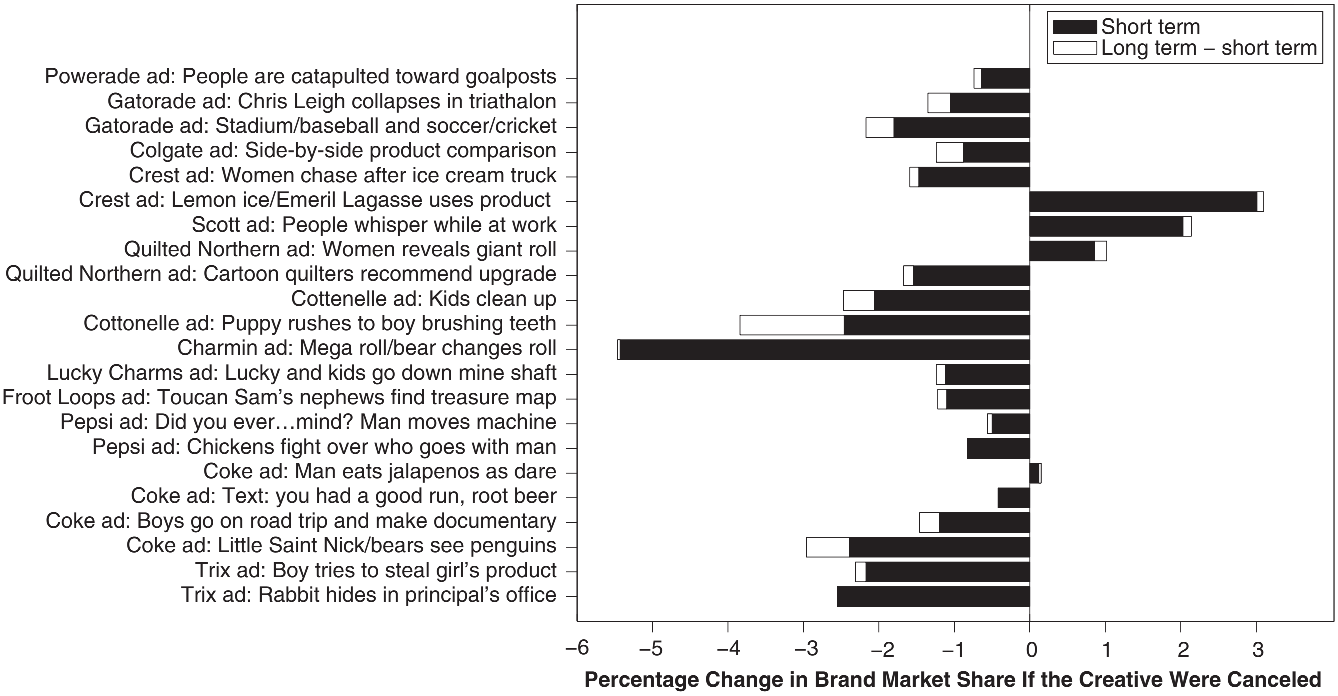
[16]Specifically, we first compute the purchase probability for each brand in the consumer's first observed shopping trip with and without an advertising creative and simulate a purchase decision based on these probabilities. Then, using the purchase event feedback term (lag brand purchase), we compute consumers' purchase probabilities in the second shopping trip. We continue this process until the last observed shopping trip, thereby obtaining a new purchase sequence for each consumer. The entire process is repeated 100 times for each consumer, and the new market share is obtained by averaging the simulated market shares obtained under the 100 simulations. The effect of advertising is then the difference between simulated long-term effects with and without an advertising creative.

[17]In Web Appendix E, we apply a "model-free" approach to measure advertising effects, which does not involve assumptions regarding functional form, advertising decay, and so on. We also find small but positive advertising effects.

[18]Unlike display advertising, in which ads and content are coincident (i.e., appear on the same page), ads and content are sequential in television advertising. Thus, it is not possible to guarantee ex ante (at the start of a pod) that a consumer will watch the subsequent ad, as would be possible for display ads.

Figure 7

PERCENTAGE CHANGE IN OWN-BRAND MARKET SHARE FOLLOWING ELIMINATION OF AN ADVERTISING CREATIVE



collecting additional sales data, one can improve the returns to advertising.

Table 5 summarizes the foregoing discussion by classifying the targeted advertising approaches into advance or real-time buy and the associated performance metrics into costs or profit. Different cells utilize different model outputs and data. The upper left cell uses the viewing data and the advertising viewing model; the upper right cell uses the viewing and sales data and the advertising viewing model and sales response model; the bottom cells use not only data and models in the respective cell above, but also the show viewing model. Thus, the bottom cells have the most substantial data and computational requirements.[19]

We explore these policies using the leading brand of bathroom tissue, Charmin, over the holdout period of July 2006.[20] Our analyses abstract away from competitive response (from the retailers, the competitors, or the networks) and assume a fixed (per-exposure) advertising price to a show. Thus, the findings are best interpreted as a marginal improvement in advertising, holding all else fixed. While this assumption is reasonable when considering changes in purchases by a single advertiser, a more systematic change in policy would require an assessment of how advertising rates might change in response to new targeting capabilities, and how the advertiser's competitors might adjust their own advertising schedules in response. While a fruitful area for additional research, this analysis extends beyond the scope of this study, and we interpret our results in light of this caveat.

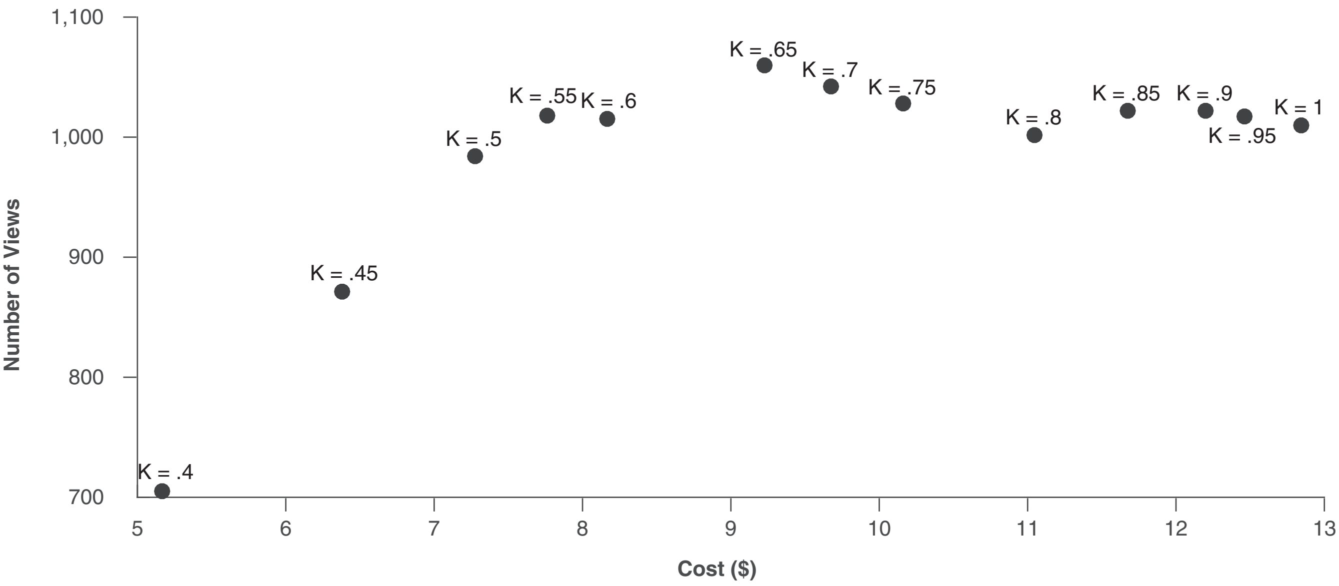
Cost-based real-time buy. We first consider the potential for Charmin to lower its per-view cost (views differ from exposures inasmuch exposures need not be viewed). To do this, we first compute each targeted household's observed average per-exposure advertising cost, by averaging the total observed advertising costs across all exposures to that household in the holdout data. Next, we compute the predicted cost per view by dividing the total observed advertising costs for a household by the predicted viewership. For example, if there are 2,000 exposures and 1,000 views to household i, and the placement costs $500, then the average cost per exposure (denoted c_i) is $500/2,000 = $.25, and average cost per view (denoted c_i^*) is $500/1,000 = $.50. Fifty parameter draws from the vector of parameter estimates are used to predict viewing probabilities.

Denote the predicted cost for a particular show's views on network n at time t to household i as c_{itn}^*, which is the ratio of the view's exposure cost and its predicted viewing probability. Intuitively, if $c_{itn}^* < c_i^*$, then a more efficient allocation of expenditures is feasible in terms of cost per view by reallocating ad dollars to show tn. Therefore, a parsimonious rule for the advertiser would be to set a level

[19]In Web Appendix F, we also consider an aggregate-level targeting approach wherein advertisers can buy only shows and not users. This corresponds to current practice. Results indicate that leveraging household viewing information substantially increases advertising efficiency even in this aggregated case.

[20]We focus on bathroom tissue because it has broad penetration. Within this category, we focus on Charmin because it has the largest number of advertising views (share = 46.8%) and the second largest sales (share = 24.1%). As a result, there is sufficient information from which to infer advertising response. In the holdout period of July 2006, the 834 sample households viewed 690 of Charmin's advertisements, at a total cost of $8.77. In the profit-based simulations, we consider Charmin's advertising creative "Mega roll/bear changes roll," which has significant effect on one consumer segment (.46; SE = .16) and insignificant effect on the other segment (.05; SE = .20). As a result of the heterogeneity in advertising response, targeting should be more efficacious.

Figure 8
ADVERTISING VIEWS AND COSTS UNDER DIFFERENT PURCHASE THRESHOLDS (*K*)



Notes: The current observed cost is $8.77 with 690 views and 780 exposures.

K, conditioned on the information available at time *t*, such that the advertiser purchases an advertising slot in show tn if $c^*_{itn} < Kc^*_i$. For example, a rule of K = 1 implies the advertiser should buy all slots with a per-view cost that is lower than the average cost per view under the current schedule. Lower values of K imply a rule that leads to higher levels of purchase efficiency, but at the cost of reach (there will be only a limited number of slots that meet an increasingly stringent cut-off). We implement this purchase rule viewer by viewer, beginning at the first advertising slot in the period and ending when either (1) the period has ended or (2) both advertising views and expenditures under the new rule exceed those observed in the data.[21]

Figure 8 portrays the total advertising views and costs (across households) under different values of K. In the figure, we note that total advertising costs decrease as K decreases. This is primarily because a lower value for K means that Charmin is buying advertisements that have lower costs per view, and because fewer advertising slots are available that meet this criterion. The effect of decreasing K on views is more complex. On the one hand, a decrease in K means fewer slots are available that meet that criteria of lower cost per view (as K goes to 0, there will be no advertising). This implies reduced views owing to reduced reach. On the other hand, there is the potential that lower cost per view can lead to more views because one can buy more views for a fixed budget (with sufficient inventory available). How these two opposing forces trade off is an empirical question.

Findings suggest that when K is set below .6 (implying high cost-efficiency constraint on ad buys), advertising views surpass those under the current schedule (690) even though overall advertising costs are below the current level ($8.77, or $12.71 CPM). For example, when K = .4, Charmin's views are increased by 2% while its costs are decreased by 41%; when K = .6, Charmin's views are increased by 47% while its costs are decreased by 7%. Charmin buys fewer advertisements than under the current schedule, but the advertisements cost less and are more likely to be seen. Thus, it is possible to lower costs *and* increase views. To maintain the same number of views (690), the cost is even lower. For instance, when K = .4, the CPM is $7.38, 41% lower than the current level; when K = .6, the CPM is $7.99, 37% lower than the current level.

Of additional interest is the decomposition of real-time buying cost-efficiency gains into (1) gains from placing advertisements in shows that have lower cost per exposure (i.e., reducing cost) and (2) gains arising from targeting advertisements to those who are less likely to avoid them (i.e., increasing views). To answer this question, we replace the model's forecasted advertising avoidance (denoted Forecasted) with the average observed advertising avoidance rate across viewers (denoted Observed); we then use Observed to reimpute advertising costs and views under the simple buying rule. Setting K = .4, we first find that is no longer the case that views increase and costs decrease under Observed (i.e., no heterogeneity in ad avoidance); instead, both increase. Second, the average costs per view in the observed data are 1.27 cents, .73 cents under Forecasted, and .94 cents under Observed. Thus, show placement alone yields a 26% improvement in cost efficiency over the observed schedule, but coupled with the advertisement viewing model, there is a 43% improvement. In other words, roughly two-fifths of the improvement in efficiency is due to reducing advertising avoidance and the other three-fifths is due to cheaper impressions.

Cost-based advance buy. In this targeting scenario, the advertiser purchases advertising slots in advance for a given period, minimizing the total cost of expected views for targeted households. In other words, advertisers are unable to condition

[21]A generalization of this purchase rule is to soften either the minimum advertising views or the maximum budget constraints. This exercise yields similar insights as described subsequently.

on current viewing, meaning that they need to predict show viewing as well as advertising viewing.[22,23]

For each household i, the advertiser selects shows that can (in expectation) maintain the total advertising views under the current schedule at the lowest cost. The optimization problem can be written as follows:

$$\text{(20)}\qquad \mathrm{Min}_{\{x_{itn}\}_{t,n}} \sum_t \sum_n c_{tn} x^*_{itn},$$

$$\text{(21)}\qquad \text{s.t.}\quad x^*_{itn} \in \{0,1\}, \forall t, n, \text{ and}$$

$$\text{(22)}\qquad E_{\Theta_i}\left(\sum_t \sum_n x^*_{itn} r_{itn}\right) \ge E_{\Theta_i}\left(\sum_t \sum_n x_{itn} r_{itn}\right),$$

where Equation 20 represents the expected cost of advertising under schedule $\{x^*_{itn}\}$. The term x^*_{itn} denotes household-show selection under the optimal schedule, where $x^*_{itn} = 1$ if show tn is selected for household i, $x^*_{itn} = 0$ otherwise. Similarly, $x_{itn} \in \{0,1\}$ denotes household-show selection under the current schedule. Finally, c_{tn} denotes the per-exposure advertising price associated with show tn and r_{itn} is the probability that household i watches the advertisement placed in show tn. The constraint in Equation 22 ensures that the firm is minimizing costs for the viewers it currently targets. The expectations in Equation 22 are taken over the distribution of r_{itn}, which is associated with the advertising viewing estimates, $\Theta_i \equiv \{\beta^s_i, \beta^O_i, \beta^A_i, \lambda_i, c_i\}$. The value of r_{itn} is obtained from the viewing model output as the product of sampling the show, watching the show, and not exiting the show before the advertisement appears; it takes into account the uncertainty in advertising location within the show:

$$\text{(23)}\qquad r_{itn} = \Pr\left(y^S_{itn} = 1\right)\Pr\left(y^W_{itn} = 1 \middle| y^S_{itn} = 1\right) \times \left(\int_{t'} \Pr\left(l^W_{itn} \ge t'\right)\Pr\left(y^A_{it'n} = 1\right) f(t')\,dt'\right),$$

where t' denotes a possible advertising location (time into show) and $f(t')$ represents the probability density function of a uniform distribution, whose support is advertising pods in show tn. Assuming zero viewing offset, conditional on watching show tn, household i will be exposed to the advertisement placed at t' if the viewing length l^W_{itn} exceeds t'. The term $\Pr(y^A_{it'n} = 1)$ denotes the probability that this advertisement will not be zapped (if the show is live) or zipped (if the show is recorded).

Because $E_{\Theta_i}(\sum_t \sum_n x^*_{tn} r_{itn}) = \sum_t \sum_n x^*_{tn} E_{\Theta_i}(r_{itn})$, we first compute $E_{\Theta_i}(r_{itn})$ by simulation[24] and then use it to solve the optimization problem for each household. The average per-view advertising price reduces from 8.0 cents to .7 cents. The overall cost reduces from \$8.77 to \$.72, a 92% reduction in expenses. The cost reduction per 1,000 households is \$9.65.

[22]Because almost all recorded shows arise from automated recording, the automated list is used to predict the shows that will be recorded.

[23]We assume no network guarantee on minimal views to advertisers as sometimes happens in practice. Web Appendix G considers an extension that includes rating guarantees, that is, a scenario wherein advertisers pay for guaranteed views instead of exposures.

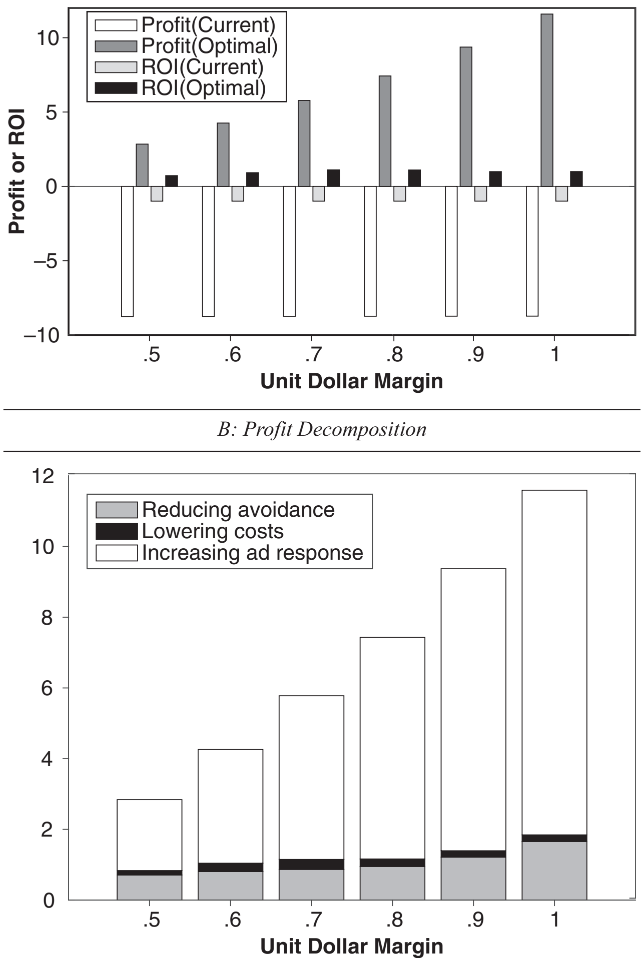
[24]Specifically, we draw 50 sets of household-specific parameter estimates, each leading to a prediction of the household's show choices and viewing length conditional on choice. The value of $E_{\Theta_i}(r_{itn})$ is taken as the average of this predicted viewership across the 50 sets of parameter draws.

Figure 9

REAL-TIME BUY SIMULATED GAINS FOR SAMPLE HOUSEHOLDS UNDER DIFFERENT UNIT DOLLAR MARGINS

A: Simulated Incremental Profit and ROI

B: Profit Decomposition



Profit-based real-time buy. To ascertain the effect of advertising on profits, we compare the marginal effect of advertising to a given household with its cost. A profitable buy is one wherein the marginal revenue exceeds the cost.

Incremental profits from advertising are computed using the estimates from the advertising response model described in the "Advertising Response" section. Following Equation 18, the effect of a view of creative a on household i's purchase probability for brand j in shopping trip m is[25]

$$\text{(24)}\qquad \Delta_{ijm} = \Pr\left(y^P_{ijm} = 1 \middle| N^A_{ijam} = 1\right) - \Pr\left(y^P_{ijm} = 1 \middle| N^A_{ijam} = 0\right)$$
$$= \frac{\exp\left(Z_{ijm}\theta_i + h_i\left(A_{ijm}\middle| N^A_{ijam} = 1\right)\right)}{1 + \sum_{j'=1}^{J}\exp\left(Z_{ij'm}\theta_i + h_i\left(A_{ij'm}\middle| N^A_{ijam} = 1\right)\right)} - \frac{\exp\left(Z_{ijm}\theta_i + h_i\left(A_{ijm}\middle| N^A_{ijam} = 0\right)\right)}{1 + \sum_{j'=1}^{J}\exp\left(Z_{ij'm}\theta_i + h_i\left(A_{ij'm}\middle| N^A_{ijam} = 0\right)\right)},$$

[25]Product category subscript is omitted for simplicity.

Table 6
TARGETING GAINS

	Cost-Based	Profit-Based
Real-time buy	Target views increase 47% and costs decrease 7%; alternatively, costs decrease 37% for same target views.	Profits per thousand increase from −\$10.4 to \$3.4–\$13.9. Normed to an ROI of −100%, this increase implies an ROI of 108%–139%.
Advance buy	Costs decrease 92% for same target views.	Negligible spending and profit.

where the first term represents sales with advertising and the second term indicates sales without advertising.

The lift yields incremental unit sales, not profits. Because we do not observe unit margins, we solve for the incremental profit under different unit dollar margins l.[26] The expected incremental profit of advertisement tn is calculated as $\Pr(y^A_{itn} = 1)\Delta_{ijm}l - c_{tn}$, where $\Pr(y^A_{itn} = 1)$ denotes the probability that the advertisement will not be avoided and c_{tn} is the cost of advertising slot tn. Advertisement tn is purchased for household i if its expected incremental profit is positive.

Figure 9, Panel A, depicts the simulated profits and ROI (defined as profit per ad dollar invested) under the observed and proposed advertising schedules for different margins. The simulated short-term advertising profit using the observed schedule is negative, with losses of about \$8.70 across the 834 households in our sample. In contrast, the proposed schedule produces a positive simulated incremental profit (the net gain ranges from \$2.84 to \$11.59). The net gain for 1,000 households ranges from \$3.41 to \$13.90.

Further insights into this gain can be obtained by decomposing three factors driving the expected advertising lift, $\Pr(y^A_{itn} = 1)\Delta_{ijm}l - c_{tn}$: (1) reducing avoidance (i.e., increasing $\Pr(y^A_{itn} = 1)$); (2) targeting people with higher advertising response (i.e., increasing Δ_{ijm}); and (3) shifting to cheaper impressions (i.e., decreasing c_{tn}). Factors 1 and 3 lower the cost of advertising, whereas factor 2 increases the revenue. To achieve this aim, we recompute profit gains under three alternative scenarios. In Model A, forecasted advertising avoidance is replaced by the average observed avoidance rate across viewers. In Model B, estimated advertising response is replaced by the average observed advertising response across viewers. In Model C, advertising cost is replaced by the average observed advertising cost across slots. The difference in profits between the constrained models A–C and the respective models where these effects vary across users yields the gains from targeting for each respective component (reducing avoidance, increasing ad response, and lowering costs). Figure 9, Panel B, depicts this profit decomposition under different unit dollar margins. Depending on margin, the majority (75%–89%) of gain comes from targeting people with higher advertising response, about 9%–19% of gain comes from reducing advertising avoidance, and the remaining portion comes from improvement in cost efficiency. The finding that the greatest increase in profits accrues to increasing revenue by targeting those most responsive to advertising suggests the value of single-source data. Combining purchase data with viewing leads to the greatest share of profit gains.

Profit-based advance buy. The profit-based advance-buy scenario requires advertisers to forecast show viewership, advertising viewership, and sales. On an intuitive level, one might expect this information requirement to lower the efficacy of advertising because more advertisements would be aired to those who would not see them. As this statistical uncertainty reduces the expected advertising response but not advertising expense, one might expect advertising to become less effective (thereby leading to lower optimal levels of advertising). More formally, as noted in the "Profit-Based Real-Time Buy" section, the expected incremental profit from an advertisement is $\Pr(y^A_{itn} = 1)\Delta_{ijm}l - c_{tn}$. With advance buying, this expected incremental profit becomes $r_{itn}\Delta_{ijm}l - c_{tn}$, with r_{itn} as defined in Equation 23. Advertisement tn is purchased for household i if its expected incremental profit is positive.

Implementing this optimization, we find overall spending reduces to zero. In expectation, most advertising is unprofitable because of relatively low advertising elasticities coupled with little attendant diminution in media costs. Hence, in contrast to real-time purchases, the optimal advertising levels decrease.

Targeting summary. The targeting results are summarized in Table 6. Results suggest there are large percentage-wise decreases in costs and increases in views possible as a result of improved targeting, and that major gains can accrue when shifting to real-time buys. Further, short-term profits can be increased by targeting people with higher advertising response, selecting advertising slots with higher viewing probability, and shifting to cheaper impressions. Of note, the allocation of surplus among multiple advertisers and TV networks in the long run depends on competition in both the product market and the advertising market, as well as the advertising pricing and allocation mechanisms. For instance, if TV networks could raise their price in the profit-based real-time buy scenario, then they could extract all surplus from the advertiser. The average per-view price increase feasible to the network is therefore obtained by dividing the profit difference between the optimal schedule and the current schedule with the number of ads under the optimal schedule. This calculation implies that prices could be 44%–68% higher than the current level.

We conclude by noting several additional caveats pertaining to the profit-based targeting approaches. First, household-level scanner data comprise only a fraction of the viewers and therefore might not be wholly representative of all consumers. Second, single-source data are costly to develop because multiple companies own different pieces of the data, making the integration of these data sources difficult. Third, the purchase data used in this study are limited to CPG, and therefore the degree to which our findings generalize to other categories remains unclear. Nonetheless, our

[26]Bathroom tissue has a gross profit margin of roughly 13%. In the data, the mean and median prices of Charmin are, respectively, \$6.80 and \$6.30, leading to a unit dollar margin of about \$.80–\$.90.

preliminary findings suggest that additional investment might be warranted to address these concerns.

CONCLUSION

TV remains by far the predominant form for the transmission and reception of video content, and the largest advertising medium. Moreover, its preeminence stands to benefit from recent digital innovations such as DVRs and STBs. Yet digital TV is both a blessing and a curse for advertisers. On the one hand, STBs have greatly enhanced viewers' TV viewing experiences, leading to increased viewing consumption and the opportunity for advertising targeting at the TV-set level. On the other hand, these boxes enable viewers to fast-forward through advertisements. It is our goal to redress this limit by taking advantage of micro-level viewing data and micro-targeting capabilities inherent in STBs to better understand viewer behavior and accordingly improve the efficacy of advertising. If targeting proves effective, so-called programmatic television opens new paths to TV advertising pricing by allowing TV networks and cable companies to sell "boxes" as well as shows to advertisers.

To better assess the nature of advertising consumption, we use set-top data to characterize viewing behavior and purchase data to link ad viewing to purchases. Viewing behaviors involve a series of conditional decisions that lead to an ad exposure: whether to watch TV and for how long, which show to watch and for how long, and whether to watch ads within those shows. Our first finding is that most viewers watch their TV most of the evening. This means variation in ad exposure is more a function of the show chosen than the decision to watch TV. Second, we find that viewers tend to sample shows before selecting them, and once a show is selected, they tend to watch to the end. This suggests that switching away from ads in live viewing contexts is not the modal behavior. However, when the selected show is recorded, it is the norm to fast-forward through advertisements on the way to the completion of the show. Combined, these two observations suggest that targeting ads to people who watch live TV leads to higher viewership. Third, we find that the preponderance of variation in ad avoidance is explained by person-level factors, suggesting that (1) gains from targeting might be considerable and (2) targeting by shows, the current norm, limits advertiser ability to prevent ad avoidance (within shows, news and reality shows exhibit the most switching, while drama exhibits the least). Last, we find advertising effects to be small and positive, consistent with previous research (Lodish et al. 1995), and we find heterogeneity in ad response across consumers. This suggests the potential to improve targeting by reaching responsive consumers.

Given the potential benefits from targeting, we consider several household-level targeting scenarios by manipulating (1) whether the objective function is to minimize costs for a given set of views or to maximize incremental profit from advertising, and (2) whether the advertising purchase is made in advance or in real time. Results indicate microtargeting can lower advertising costs and raise incremental profit even in the face of ad avoidance. We find that the greatest potential to increase the profitability of advertising arises from (1) the integration of purchase data with viewing data and (2) the ability to buy placements in real time instead of in advance. As for the former, single-source data are becoming increasingly available and purveyed by firms such as Acxiom TV. As for the latter, there is a developing advertising ecosystem that could enable real-time buying, as is commonly observed in display advertising. Our analysis suggests the importance of these technological advances.

Several extensions are possible. First, the sales model does not consider context effects in advertising. As a result, the proposed targeting strategies do not account for context effects. Consumer behavior research suggests that advertisements placed within a program of dissimilar content are recalled significantly better than those placed within a program of similar content (e.g., Furnham and Price 2006). One extension is to incorporate context effects in measuring advertising effectiveness and apply it in designing targeting strategies.

Second, the targeting strategies we propose do not consider potential complementarities between the consumption of goods and advertisements (Becker and Murphy 1993). According to a recent study by Tuchman, Nair, and Gardete (2017), the quantity of the advertised product purchased recently can explain advertising avoidance rates. These authors suggest that advertising efficacy depends on the compliance of the targeted households and that the firm can target the subset of households whose purchases and welfare will likely change in response to the advertising campaign.

Third, the targeting strategies are designed according to an integrated channel structure, under which advertisers and TV networks share the same information and same objectives. Therefore, a limitation is that we are unable to model the pricing mechanism and the advertising allocation problem from the TV networks' perspective. We use this design because advertising allocation can be designed to align networks' and advertisers' incentives (e.g., Wilbur, Xu, and Kempe 2013). Still, another extension is to consider strategic interactions between advertisers and TV networks and how they affect gains available from microtargeting.

Fourth, with minor modifications, our model can also be applied in various other contexts, such as media consumption in an online environment and with mobile devices. For example, in such contexts, viewers in general face a larger choice set because they can sample shows whenever they want. In addition, some tablet options do not allow viewers to avoid advertisements, so the viewing behavior could be modeled by eliminating advertising viewing decision.

Finally, future research can extend microtargeting by taking into account competitive response and allowing reoptimization of advertising and product prices. Owing to the growth in digital TV, we believe these and other extensions will yield economically consequential insights in the coming years.

REFERENCES

Anand, Bharat N., and Ron Shachar (2009), "Targeted Advertising as a Signal," *Quantitative Marketing and Economics*, 7 (3), 237–66.

Anand, Bharat N., and Ron Shachar (2011), "Advertising, the Matchmaker," *RAND Journal of Economics*, 42 (2), 205–45.

Arcidiacono, Peter, Patrick Bayer, Jason R. Blevins, and Paul B. Ellickson (2015), "Estimation of Dynamic Discrete Choice Models in Continuous Time with an Application to Retail Competition," *Review of Economic Studies*, 83 (3), 889–931.

Becker, Gary S., and Kevin M. Murphy (1993), "A Simple Theory of Advertising as a Good or Bad," *Quarterly Journal of Economics*, 108 (4), 941–64.

Bronnenberg, Bart J., Jean-Pierre Dubé, and Carl F. Mela (2010), "Do Digital Video Recorders Influence Sales?" *Journal of Marketing Research*, 47 (6), 998–1010.

Bruell, Alexandra (2017), "TV's Next Act: Targeting Ads at Yogurt Lovers and Home Buyers," *Wall Street Journal* (June 18), https://www.wsj.com/articles/tvs-next-act-targeting-ads-at-yogurt-lovers-and-home-buyers-1497778200.

Eastlack, Joseph O., and Ambar G. Rao (1989), "Advertising Experiments at the Campbell Soup Company," *Marketing Science*, 8 (1), 57–71.

eMarketer (2016), "US Digital Ad Spending to Surpass TV This Year," eMarketer.com (September 13), https://www.emarketer.com/Article/US-Digital-Ad-Spending-Surpass-TV-this-Year/1014469.

Esteves-Sorenson, Constanca, and Fabrizio Perretti (2012), "Micro-Costs: Inertia in Television Viewing," *Economic Journal*, 122 (563), 867–902.

Furnham, Adrian, and Marie-Therese Price (2006), "Memory for Televised Advertisements as a Function of Program Context, Viewer-Involvement, and Gender," *Communications*, 31 (2), 155–72.

Gal-Or, Esther, Mordechai Gal-Or, Jerrold H. May, and William E. Spangler (2006), "Targeted Advertising Strategies on Television.," *Management Science*, 52 (5), 713–25.

Goettler, Ronald L. (2012), "Advertising Rates, Audience Composition, and Competition in the Network Television Industry," working paper, University of Rochester, http://goettler.simon.rochester.edu/research/papers/adrates.pdf.

Gustafson, Paul, and S. Siddarth (2007), "Describing the Dynamics of Attention to TV Commercials: A Hierarchical Bayes Analysis of the Time to Zap an Ad," *Journal of Applied Statistics*, 34 (5/6), 585–609.

Hu, Ye, Leonard M. Lodish, and Abba M. Krieger (2007), "An Analysis of Real World TV Advertising Tests: A 15-Year Update," *Journal of Advertising Research*, 47 (3), 341–53.

Kamakura, Wagner A., and Gary J. Russell (1989), "A Probabilistic Choice Model for Market Segmentation and Elasticity Structure," *Journal of Marketing Research*, 26 (4), 379–90.

Kim, Jun B., Paulo Albuquerque, and Bart J. Bronnenberg (2010), "Online Demand Under Limited Consumer Search," *Marketing Science*, 29 (6), 1001–23.

Lee, Mei-Ling Ting, and G.A. Whitmore. (2006), "Threshold Regression for Survival Analysis: Modeling Event Times by a Stochastic Process Reaching a Boundary," *Statistical Science*, 21 (4), 501–13.

Lehmann, Donald R. (1971), "Television Show Preference: Application of a Choice Model," *Journal of Marketing Research*, 8 (1), 47–55.

Lodish, Leonard M., Magid Abraham, Stuart Kalmenson, Jeanne Livelsberger, Beth Lubetkin, Bruce Richardson, et al. (1995), "How TV Advertising Works: A Meta-Analysis of 389 Real World Split Cable TV Advertising Experiments," *Journal of Marketing Research*, 32 (2), 125–39.

Lovett, Mitchell, and Michael Peress (2015), "Targeting Political Advertising on Television," *Quarterly Journal of Political Science*, 10 (3), 391–432.

Malaviya, Prashant, Jolita Kisielius, and Brian Sternthal (1996), "The Effect of Type of Elaboration on Advertisement Processing and Judgment," *Journal of Marketing Research*, 33 (4), 410–21.

Moshkin, Nickolay V., and Ron Shachar (2002), "The Asymmetric Information Model of State Dependence," *Marketing Science*, 21 (4), 435–54.

Nevskaya, Yulia, and Paulo Albuquerque (2012), "A Continuous Time Model of Product Usage: Measuring the Effect of Product Design and Rewards in Online Games," working paper, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2098915

Nielsen (2016), "The Nielsen Total Audience Report: Q2 2016," http://www.nielsen.com/us/en/insights/reports/2016/the-nielsen-total-audience-report-q2-2016.html.

Rust, Roland T., and Mark I. Alpert (1984), "An Audience Flow Model of Television Viewing Choice," *Marketing Science*, 3 (2), 113–24.

Schweidel, David A., and Robert J. Kent (2010), "Predictors of the Gap Between Program and Commercial Audiences: An Investigation Using Live Tuning Data," *Journal of Marketing*, 74 (3), 18–33.

Sethuraman, Raj, Gerard J. Tellis, and Richard A. Briesch (2011), "How Well Does Advertising Work? Generalizations from Meta-Analysis of Brand Advertising Elasticities," *Journal of Marketing Research*, 48 (3), 457–71.

Shachar, Ron, and John W. Emerson (2000), "Cast Demographics, Unobserved Segments, and Heterogeneous Switching Costs in a television Viewing Choice Model," *Journal of Marketing Research*, 37 (2), 173–86.

Siddarth, S., and Amitava Chattopadhyay (1998), "To Zap or Not to Zap: A Study of the Determinants of Channel Switching During Commercials," *Marketing Science*, 17 (2), 124–38.

Tellis, Gerard J., Rajesh K. Chandy, Deborah MacInnis, and Pattana Thaivanich (2005), "Modeling the Microeffects of Television Advertising: Which Ad Works, When, Where, for How Long, and Why?" *Marketing Science*, 24 (3), 351–66.

Tuchman, Anna, Harikesh S. Nair, and Pedro Gardete (2017), "TV Ad Skipping, Consumption Complementarities and the Consumer Demand for Advertising," *Quantitative Marketing and Economics*, published electronically October 2017, https://doi.org/10.1007/s11129-017-9192-y.

Urbanski, Al (2016), "Addressable TV Isn't on the Way; It's Arrived," *Direct Marketing News* (February 25), http://www.dmnews.com/marketing-strategy/addressable-tv-isnt-on-the-way-its-arrived/article/478386/.

Weitzman, Martin L. (1979), "Optimal Search for the Best Alternative," *Econometrica*, 47 (3), 641–54.

Wilbur, Kenneth C. (2008), "A Two-Sided, Empirical Model of Television Advertising and Viewing Markets," *Marketing Science*, 27 (3), 356–78.

Wilbur, Kenneth C., Linli Xu, and David Kempe (2013), "Correcting Audience Externalities in Television Advertising," *Marketing Science*, 32 (6), 892–912.

Yang, Sha, Vishal Narayan, and Henry Assael (2006), "Estimating the Interdependence of Television Program Viewership Between Spouses: A Bayesian Simultaneous Equation Model," *Marketing Science*, 25 (4), 336–49.