TV Viewing and Advertising Targeting

Yiting Deng                  Carl F. Mela *

September 6, 2015

*Yiting Deng is an Assistant Professor at the Mendoza College of Business, University of Notre Dame (email: ydeng@nd.edu, phone 510-735-6255). Carl F. Mela is the T. Austin Finch Foundation Professor of Business Administration at the Fuqua School of Business, Duke University (email: mela@duke.edu, phone 919-660-7767). The authors would like to thank Bryan Bollinger, Pedro Gardete, Roni Shachar, Richard Staelin, Peng Sun, Kenneth Wilbur, Yi (Daniel) Xu, and seminar participants at Duke University, University of California-Riverside, Erasmus University, HKUST, McGill 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. The usual disclaimer applies.
Abstract

Television (TV), the predominant advertising medium, is being transformed by the micro-targeting capabilities of digital video recorders (DVRs) and set-top boxes (STBs). Though micro-targeting is common on the Internet, TV represents a more substantial share of consumer time and attention, suggesting even greater potential for micro-targeting communications. Accordingly, this paper uses a proprietary, household-level, single source data set to develop a second-by-second show and advertisement viewing model, using this approach to forecast consumers’ exposure to advertising and the downstream consequences for sales.

We find that micro-targeting simultaneously lowers advertising costs and increases advertising views among brands’ currently targeted consumers, and that these advantages are amplified when advertisers are allowed to buy real-time as opposed to up-front. However, most advertisements are not profitable in the short-term, suggesting that an advertisement schedule that maximizes profits in the short-run typically involves fewer advertisements than those observed in the data. Overall, we find considerable targeting gains.

Keywords: TV advertising, targeting, DVR, sampling

JEL Classification Codes: M31, M37, L10, L82, C61

1 Introduction

Television (TV) is the most prominent modality for the transmission and reception of video content. According to Nielsen’s total audience report (Nielsen (2015)), American adults spend about 36 hours watching traditional TV each week. In contrast, the weekly average time spent on PC, smartphone and tablet is about 16 hours. TV also remains, by far, the largest advertising medium. Bureau (2015) reports that $65.7 billion was be spent on TV advertising in U.S. in 2014, accounting for 38% of total U.S. advertising spending, while Internet advertising accounts for 29% of total U.S. advertising spending, though it is expected to overtake TV advertising in 2018 (eMarketer (2014)).

With the proliferation of digital set-top boxes (STBs) and video recorders (DVRs), TV is likely to remain preeminent. 58% of households with HDTV owned DVRs as of February 2014 and many others own STBs (Nielsen (2014)). STBs and DVRs are transformative in two regards. First, digital TV affords household-level measures of TV viewing, making it possible to better forecast viewer advertising
exposure. Second, it enables household-show level targeting, enhancing targeting precision. Though uncommon in TV, this type of micro-targeting has been extensively applied in online advertising. Targeting tools have been developed by Google, Yahoo!, Facebook, YouTube, Yelp and a number of other Internet companies/website. Micro-targeting on the Internet is highly successful, and can likewise be applied to TV.

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 campaigns at the household-show level. Several companies have already started offering these types of targeted TV advertising solutions to advertisers. For instance, TiVo Research and Analytics, Inc. (TRA) developed a software platform “Media TRAnalytics®”, which combines household-level TV tuning and purchase data to help advertisers achieve higher ROI. Similarly, Nielsen Catalina Solutions launched “AdVantics on Demand (TM)” that helps advertisers achieve better targeting based on retail purchase data and NBCUniversal announced the launch of its Audience Targeting Platform (ATP) that will use viewing and purchase histories to identify client-specific inventory across NBCUniversal’s portfolio of national broadcast and cable networks. In light of these evolving advances in digital TV, it is our goal to use set-top data to model households’ second-by-second TV and advertising consumption behavior and integrate the model with purchase data in order to propose ways for advertisers to improve their targeting.

With regard to TV and advertising consumption, we begin by documenting a number of stylized facts about viewing in order to inform our modeling choices. For example, households watch prime-time TV almost every night (mode = 95% of days) and for most of the evening (mode = 100% of prime-time hours). Once a viewing session commences, it is commonplace for a viewer to sample several shows prior to selecting one to view. Once a show is selected, it is commonplace for a viewer to watch a show to its conclusion. Within a show, we find that viewers’ advertising avoidance is more common when the show is recorded. Person-specific factors explain most of the variation in advertising avoidance (20.4%). The person-specific variation dwarfs the variation explained by genre (0.7%), brand (0.3%), time (0.0%), and category (0.0%). This variance decomposition suggests the potential for significant

1 http://advertising.yahoo.com/targeting/
2 http://www.facebook.com/help/131834970288134/
3 http://www.youtube.com/yt/advertise/targeting.html
4 http://www.yelp.com/advertise/agency/targeting
returns to household-level advertising targeting, and that targeting on genres and time will have relatively little effect 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. Factors playing a major role in show consumption utility include its length, genre, network, familiarity, and offset. When consumers encounter an advertisement in these shows, we again draw upon this concept of flow 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, and the utility is higher in the first slot of a commercial break.

With regard to targeting, we consider: 1) whether an advertiser seeks to minimize the costs of its target exposures or to maximize its incremental profit from advertising; and 2) whether the advertising purchase is made in advance or in real-time. Advance purchase, or “up-front,” is the current norm in TV advertising sales, but with the increased potential for firms to buy advertising real-time, like on the Internet, the real-time buying is becoming increasingly relevant. With advance buy, we find that it is possible to lower costs per target exposure by over 80%. With real time buy, it is further possible to lower target costs per exposure while concurrently increasing target exposures; in one schedule exposures to the target households can be increased by 47% while concurrently reducing costs by 12%. In short, we document dramatic increases in the cost efficiency of targeted media buys. Likewise, we find that advertisers can improve their short-run advertising profitability, but this improvement arises primarily by reducing advertising spend. Most advertisements exhibit low sales response and are therefore unprofitable in the short run. This suggests one would need to posit long-term branding effects of advertising to rationalize TV advertising spend.

The remainder of the paper is organized as follows. First, §2 reviews the relevant literature. §3 then describes TV viewing behavior to motivate the viewing model presented in §4. §5 discusses estimation and identification, and §6 describes the estimation results. Based on these results and a purchase model, §7 conducts counterfactual policy experiments and evaluates potential gain to be realized from targeting. Finally, we conclude with a schedule of next steps in §8.
2 Relevant Literature

Since Lehmann (1971)'s seminal work, a rich body of literature has identified various factors affecting viewers’ utility from watching TV programs. Such factors include viewer demographics, program genre, cast demographics, advertising time, viewer’s previous program choices, and spouse’s choice (Rust and Alpert (1984); Rust et al. (1992); Shachar and Emerson (2000); Goettler and Shachar (2001); Moshkin and Shachar (2002); Yang et al. (2006); Wilbur (2008); Anand and Shachar (2011); Esteves-Sorenson and Perretti (2012)). Collectively, this line of literature suggests that person, show and time factors explain substantial variation in show viewing. Likewise, recent work suggests uncertainty pertaining to show quality affects TV show choices (Moshkin and Shachar (2002), Anand and Shachar (2011), Esteves-Sorenson and Perretti (2012) and Yao et al. (2015)). 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 looks into 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 and the media weight of a campaign (i.e., the number of times that a household had previously been exposed to a commercial) (Siddarth and Chattopadhyay (1998); Gustafson and Siddarth (2007)). Identified ad-specific factors include the frequency of the commercial, length and content of the commercial, program genre, commercial location, as well as the congruity between the commercial and the program (Norris and Colman (1993); Siddarth and Chattopadhyay (1998); Furnham et al. (2002); Furnham and Price (2006); Moore et al. (2005); Gustafson and Siddarth (2007); Teixeira et al. (2010); Schweidel and Kent (2010)).

Finally, given our goal to explore the potential of micro-targeting in TV advertising, this work also relates to the growing literature in advertising targeting. A number of papers have examined why and how targeted TV advertising works from a theoretical point of view (e.g., Gal-Or et al. (2006); Anand and Shachar (2009); Ghosh and Stock (2010)). A few papers address the issue of geographically or demographically targeted TV advertising from an empirical point of view (e.g., Kitts et al. (2010); Anand and Shachar (2011); Lovett and Peress (2014)). Our focus is instead targeting at more granular levels,
i.e., 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 Tuchman et al. (2015), who explore implications of individual-level targeting to consumers who are less likely to skip the targeted advertisements and have positive marginal advertising effect on purchase. One key point of departure in our analysis is that we model whether the show is viewed at the second-by-second level, which enables us to forecast whether the advertisement is viewed. 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.

3 Data

Micro-targeting 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 (§3.1). Next, we describe the TV program viewing data and the advertising viewing data to generate insights regarding household viewing behavior (§3.2) that help to form the basis of our viewing model.

3.1 Data Description

Several sets of data are integrated for this study, including DVR usage data, purchase data, advertising data, and programming data. This combination is often called single source data because it covers the entire TV viewing and purchase experience for a set of households. The DVR data (TiVo log files) track each household’s complete usage of a TiVo DVR 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. The TNS data are supplemented with national viewing data from AC Nielsen in order 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 DVR data, advertising data and programming data will be used to estimate
the viewing model (§4-§6). The purchase data and the Nielsen data will be used along with viewing model estimates in policy experiments on targeting (§7). We describe each dataset in turn.

3.1.1 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. The TiVo log files track each household’s moment-by-moment usage of a TiVo DVR. They record every keystroke of the DVR 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 of July 2005 - July 2006, keeping the 834 households that have both viewing and purchase information.

3.1.2 IRI DataSets (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 of June 2005 - June 2006.

The first component, the IRI purchase panel data contain 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 record panelists’ store visits. Organized by panelist-transaction time, the data include store visited and total amount spent. Combined with the purchase panel data, these store visit data enable us to infer non-purchases, defined as no purchase in a category on a given store visit.

The third component, the IRI store causal data report 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.

For a more detailed description of the field study and the data, please see Bronnenberg et al. (2010). The starting and ending date of the IRI datasets are both earlier than those respective dates for the three datasets related to TV viewing. All datasets intersect during the period of 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 hold-out validation and policy experiments.
3.1.3 TNS Advertising Schedule Data (Advertising Exposures 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, a brief description of the advertisement, 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 within the break (i.e., pod location or slot), and the estimated price of the advertisement. We infer advertising exposures by noting the time and channel of the advertisement, and assessing whether or not the channel was viewed at that time.

3.1.4 Nielsen TV Viewing Data

Because advertising rates furnished by TNS are at the show level, they do not yield a per-impression cost. As micro-targeting is at the impression level, we need to translate the show cost to an exposure cost, and do so 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 the Broadcasting & Cable magazine and report the audience size of top TV programs on broadcasting networks. We collect these data in July 2006, the period during which the policy experiments are run. These data include 376 shows on 4 networks: ABC, CBS, NBC and FOX.

3.1.5 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 (6 broadcast networks and 21 cable networks). Each view 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.

3.2 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 decisions (watch TV, sample shows, watch or record show, watch advertisements, exit show/exit TV).

3.2.1 Watching TV

Owing to the observation that most viewing (and advertising spending) takes place in prime time (defined as 8 p.m. - midnight in this paper), our ensuing analyses focus upon this daypart. Figure 1 shows most households watch TV most evenings: on average, households watch prime-time TV on 85% of the days in the sample. Figure 2 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.

Figure 1: Fraction of Days With 8 p.m.-Midnight Viewing (by Household). The x-axis is the fraction of days with 8 p.m.-midnight viewing, and the y-axis is the percent of the particular fraction in all observations.
3.2.2 Show Sampling and Viewing

When starting to watch TV, a household first chooses which show to watch.\footnote{The set of available shows includes live broadcast and an inventory of recorded shows. Patterns related to show recording will be examined in §3.2.3.} Due to incomplete knowledge of program and episode quality, households sample shows (either live or recorded) for a brief duration to decide whether to watch (Cha et al. (2008), 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 the “show completion rate” as the ratio of the time a show is actually viewed relative to its total broadcast length. Panel (a) of Figure 3 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 non-views and the completed views (the endpoints in panel (a)) enables us to zoom in on the sampling behavior (panel (b)). It suggests a power law with most sampling not exceeding a few minutes per show.\footnote{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.}

Figure 2: Frequency of Cumulative Daily TV Usage, 8 p.m.-Midnight (by Household-Day). The x-axis is the number of hours viewing, and the y-axis is the percent of the particular viewing time in all observations.
Figure 3: Show Completion Rate (by Household-Show). The x-axis is the show completion rate, and the y-axis is the percent of the particular completion rate in all observations. Panel (a) is based on all shows watched during 8 p.m.-midnight, and panel (b) excludes shows with a completion rate below 1% or above 99%.

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 to be 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 4). The hazard rate decreases drastically within the first 3 minutes, and remains relatively stable afterwards. 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 through the subsequent paper). Based on this categorization, Figure 5 indicates that, in 70% of the cases, a household decides to watch the first show sampled. Hence, sampling is informative of preferences.
In sum, the data suggest that the frequency of prime-time viewing is high, that households spend considerable time watching in the evening, and 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.
3.2.3 Show Recording

Households can choose to 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 DVRs’ storage capacity was able to store 40 hours of programming. DVR program inventories are near capacity (>90%) on most (82%) household-day pairs in the data, implying households usually have to delete one show before recording another. These recording and deletion decisions are informative about show preferences.

3.2.4 Advertising Viewing

Advertising viewership is predicated on the series of decisions shown in Figure 6. First, advertising exposure requires one to be watching the show when the advertisement airs (expose). Second, a household must decide whether to watch a show live or recorded. As most viewing is live, 78% of all advertising exposures are live. Third, when confronted with an advertisement, a household can decide to avoid it. Complete skipping occurs when a fast-forward starts before the advertisement and ends after it. Partial skipping 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). As expected, recorded shows are more subject to skipping, as evidenced by the nearly 80% skipping rate (mostly zipping) for recorded shows and a lower than 15% skipping rate (mostly zapping) for live or near-live shows.

Zapping can also be done by other means to avoid paying attention, such as leaving the room. However, as in most previous studies, we are unable to observe such behavior, and hence focus only on channel switching.
To further exemplify skipping patterns in recorded shows, we depict in Figure 7 the timing of zips and commercial breaks during a one-hour-long recorded episode of “CSI: Crime Scene Investigation” on December 8, 2005. Figure 7 indicates zips coincide with commercial breaks. The figure suggests that advertising avoidance is not uncommon, but that when watching recorded shows, viewership tends to return to the same level after the commercial break.

---

10Zipping 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 et al. (2013)).
To ascertain the factors that explain variation in prime-time advertising skipping 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 week, hour, and past skipping (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 avoidance. If, in contrast, there remains substantial household-specific variation, then the efficacy of micro-targeting 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.

Figure 7: Number of People Zipping (by Second) in One Episode of CSI
Table 1: Analysis of Variance for Advertising Skipping (Recorded Shows Watched During 8 p.m.-Midnight)

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type I SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
<th>% Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Household</td>
<td>777</td>
<td>113065.5</td>
<td>145.5</td>
<td>1160.6</td>
<td>&lt;.0001</td>
<td>20.4%</td>
</tr>
<tr>
<td>Brand</td>
<td>1920</td>
<td>1586.4</td>
<td>0.8</td>
<td>6.6</td>
<td>&lt;.0001</td>
<td>0.3%</td>
</tr>
<tr>
<td>Genre</td>
<td>14</td>
<td>4135.9</td>
<td>295.4</td>
<td>2356.2</td>
<td>&lt;.0001</td>
<td>0.7%</td>
</tr>
<tr>
<td>Network</td>
<td>55</td>
<td>2129.1</td>
<td>38.7</td>
<td>308.7</td>
<td>&lt;.0001</td>
<td>0.4%</td>
</tr>
<tr>
<td>Product category</td>
<td>573</td>
<td>168.6</td>
<td>0.3</td>
<td>2.4</td>
<td>&lt;.0001</td>
<td>0.0%</td>
</tr>
<tr>
<td>Pod</td>
<td>27</td>
<td>3964.1</td>
<td>146.8</td>
<td>1171.0</td>
<td>&lt;.0001</td>
<td>0.7%</td>
</tr>
<tr>
<td>Pod position (Slot)</td>
<td>33</td>
<td>415.7</td>
<td>12.6</td>
<td>100.5</td>
<td>&lt;.0001</td>
<td>0.1%</td>
</tr>
<tr>
<td>Day of week</td>
<td>6</td>
<td>83.8</td>
<td>14.0</td>
<td>111.5</td>
<td>&lt;.0001</td>
<td>0.0%</td>
</tr>
<tr>
<td>Hour</td>
<td>3</td>
<td>114.0</td>
<td>38.0</td>
<td>303.0</td>
<td>&lt;.0001</td>
<td>0.0%</td>
</tr>
<tr>
<td>Previous ad skipped</td>
<td>1</td>
<td>65691.3</td>
<td>65691.3</td>
<td>523923.0</td>
<td>&lt;.0001</td>
<td>11.9%</td>
</tr>
</tbody>
</table>

3.2.5 Advertising Costs

Table 2 and Figure 8 respectively provide summary statistics and distribution of per-exposure price for 15-second advertising slots for the shows discussed in §3.1.4. The median per-exposure price is about 1 cent for ABC, CBS and NBC, and is slightly above 1 cent for FOX. There also exists moderate price variation within each network.

Table 2: Summary Statistics of Per-Exposure Advertising Price Based on Nielsen Ratings (July 2006)

<table>
<thead>
<tr>
<th>Network</th>
<th>Number of Observations</th>
<th>Mean</th>
<th>Median</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC</td>
<td>94</td>
<td>0.01</td>
<td>0.01</td>
<td>0.005</td>
</tr>
<tr>
<td>CBS</td>
<td>102</td>
<td>0.012</td>
<td>0.011</td>
<td>0.004</td>
</tr>
<tr>
<td>NBC</td>
<td>93</td>
<td>0.011</td>
<td>0.010</td>
<td>0.006</td>
</tr>
<tr>
<td>FOX</td>
<td>87</td>
<td>0.015</td>
<td>0.014</td>
<td>0.010</td>
</tr>
</tbody>
</table>
To ascertain whether advertising prices relate to viewership, we regress the advertisements’ prices on ratings and network dummies. Results indicate a positive and significant relationship between price and rating, and the network coefficient is the highest for FOX, followed in turn by CBS, NBC and ABC. Presumably the differences across networks relate to the demographics of the shows viewed (e.g., Goettler (2012)). Variation in prices of exposures suggests potential efficiencies from advertising reallocation.

4 TV Viewing Model

This section describes the TV viewing model, comprised of three components motivated by the preceding empirical discussion: TV show sampling and watching, TV show recording, and advertising viewing. All three components are predicated upon the theoretical concept of flow utility, that is, the moment-by-moment consumption benefit a household derives from watching a TV show or advertisement, and non-TV activities (the outside good). In this flow utility framework, consumers derive utility from viewing a show, but 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, consumers might be uncertain about the storyline, 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, the flow utility of the show changes
(for example a new scene or show segment), and the consumer evaluates whether to continue watching
the show or 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
episodes conclude at the end of the evening, or when the consumption utility falls short of the outside
good.

We first introduce the three flow utilities in §4.1. Based on these flow utilities, we then describe the
household choice process (show sampling, show viewing, show recording, and advertising viewing) in
§4.2, §4.3 and §4.4 that follow.

4.1 Flow Utilities

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

4.1.1 Show Utility

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

\[
\begin{align*}
\bar{u}^S_{itn} &= X^S_{itn} \beta^S_i + \nu^S_{itn} + \epsilon^S_{itn} \\
&= \bar{u}^S_{itn} + \nu^S_{itn} + \epsilon^S_{itn},
\end{align*}
\]

where \( X^S_{itn} \) is a vector that captures show characteristics and household \( i \)’s past viewing behavior, including
genre, network, show length, the number of previous episodes of the program that the household has
sampled in the preceding week, and the percent of show aired when sampling begins (i.e., viewing offset).
\( \nu^S_{itn} \) represents a household-show specific error term observed by the household but not by the researcher.
\( \epsilon^S_{itn} \) represents household uncertainty pertaining to program and episode quality prior to sampling that is
revealed to the household only after sampling the show. We assume both \( \nu^S_{itn} \) and \( \epsilon^S_{itn} \) are i.i.d. standard
Type I Extreme Value distributed.
4.1.2 Advertisement Utility

Each advertisement insertion can also be represented by a unique combination of $t$ (time) and $n$ (network). The flow utility that household $i$ obtains from advertisement $tn$ is a function of the characteristics of the advertisement, $X_{itn}^A$, and is given by:

$$u_{itn}^A = X_{itn}^A \beta_i^A + \varepsilon_{itn}^A \quad (2)$$

$$\equiv \bar{u}_{itn}^A + \varepsilon_{itn}^A,$$

where $X_{itn}^A$ includes: pod, pod position (slot), genre of the associated show, product category, and whether the preceding advertisement is skipped. $\varepsilon_{itn}^A$ is an idiosyncratic error term affecting the inherent valuation of advertisement $tn$ (conditional on exposure), and it is observed by the household but not by the researcher. We assume $\varepsilon_{itn}^A$ to be i.i.d. standard Type I Extreme Value distributed.

4.1.3 Outside Good Utility

When allocating time, the household 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 household will watch TV. As such, the flow utility of the outside good is tantamount to the “opportunity cost” of time, posited to vary by household ($i$) and time ($t$), and denoted as:

$$u_{it}^O = X_{it}^O \beta_i^O + \nu_{it}^O \quad (3)$$

$$\equiv \bar{u}_{it}^O + \nu_{it}^O,$$

where $X_{it}^O$ is a vector of observable day characteristics specific to household $i$ and time $t$. These characteristics include weekday fixed effects and indicators for previous day TV viewing and previous weekday TV viewing. Month fixed effects are added to control for seasonality. $\nu_{it}^O$ is an idiosyncratic error term affecting the utility from the outside good at time $t$, and is observed by the household but not by the researcher. We assume $\nu_{it}^O$ to be i.i.d. standard Type I Extreme Value distributed.
### 4.2 TV Show Viewing

Owing to $\varepsilon_{itn}^S$ in Equation (1), households face *ex-ante* uncertainty in the utility of viewing that can only be resolved by sampling a show (that is, a brief viewing of the show). We assume that a household first samples the alternative with the highest *ex-ante* expected viewing utility given the observed show characteristics and $v_{itn}^S$, the shock unobserved to the researcher. This ordering of alternatives bears low cognitive cost and, in the absence of learning, is also consistent with the optimal search ordering implied by Weitzman (1979) and Kim et al. (2010).\footnote{Our sampling model implies households choose between the current sampled show and the remaining shows that have yet to be sampled. Thus, the consumer always chooses the last show sampled in our model. This specification is motivated by our observation that consumers rarely, if ever, return to watch previously sampled shows (occurring in only 4\% of the shows selected).} After the short exposure to the show, the household observes its utility shock for that show, $\varepsilon_{itn}^S$, and makes the decision of whether to continue watching by comparing the flow utility of this show with the expected highest flow utility to be obtained from other shows available at that time (for which the $\varepsilon_{in'n}^S$ at the other shows $n'$ are still unknown to the consumer). If the flow utility of the current show is lower, the household samples other shows.

If the flow utility of the current show is higher than the expected best remaining alternative, the household 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 $\varepsilon_{itn}^S$. 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 household compares this new utility to 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 household switches.

We elaborate on the sampling and viewing processes next.

#### 4.2.1 Sampling

Households 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 household starts by sampling the show with the highest expected flow utility. The expected flow utility of show $tn$ is $\bar{u}_{itn}^S + v_{itn}^S + E(\varepsilon_{itn}^S)$. As the $\varepsilon_{itn}^S (\forall n)$ are i.i.d. distributed,
$E(ε_{itn}^S)$ is equal across shows and the sampling decision is therefore incumbent upon $\bar{u}_{itn}^S + ν_{itn}^S$. Show $tn$ is sampled if $\bar{u}_{itn}^S + ν_{itn}^S > \bar{u}_{itn'}^S + ν_{itn'}^S$, $∀n'$ and $\bar{u}_{itn}^S + ν_{itn}^S > \bar{u}_{it}^O + ν_{it}^O$. Under the assumption that $ν_{itn}^S (∀n)$ and $ν_{it}^O$ in Equations (1) and (3) are i.i.d. standard Type I Extreme Value distributed, this probability is given by:

$$Pr\left(\gamma_{itn}^S = 1\right) = \frac{\exp(\bar{u}_{itn}^S)}{\exp(\bar{u}_{it}^O) + \sum_{n'} \exp(\bar{u}_{itn'}^S)} \quad (4)$$

where $γ_{itn}^S$ is an indicator that show $tn$ is sampled by household $i$.

After sampling show $tn$, household $i$ observes $ε_{itn}^S$ and therefore $u_{itn}^S$. The household then compares $u_{itn}^S$ with the expected highest flow utility (i.e., the inclusive value) to be obtained from the remaining non-sampled shows available at that time and the outside good:

$$IV_{it,N_t} \mid \{v_{itn}^S\}_{n}^t, ν_{it}^O = E\{ε_{itn}^S\}_{n' \in N_t} \max_{n' \in N_t} \left\{\bar{u}_{itn'}^S + ν_{itn'}^S + ε_{itn'}^S, \bar{u}_{it}^O + ν_{it}^O\right\}$$

$$= \max\left\{\bar{u}_{it}^O + ν_{it}^O, \ln\left(\sum_{n' \in N_t} \exp(\bar{u}_{itn'}^S + ν_{itn'}^S)\right)\right\},$$

where $N_t$ denotes the set of networks at time $t$ that have yet to be sampled.

If $u_{itn} ≥ IV_{it,N_t}$, household $i$ watches the show. The probability of this event is given by:

$$Pr\left(γ_{itn}^W = 1 \mid γ_{itn}^S = 1, \{v_{itn}^S\}_{n}^t, ν_{it}^O\right) \quad (5)$$

$$= Pr\left(u_{itn}^S ≥ IV_{it,N_t} \mid \{v_{itn}^S\}_{n}^t, ν_{it}^O\right)$$

$$= 1 - F_{ε_{itn}^S}(\max\left\{\bar{u}_{it}^O + ν_{it}^O, \ln\left(\sum_{n' \in N_t} \exp(\bar{u}_{itn'}^S + ν_{itn'}^S)\right)\right\} - \bar{u}_{itn}^S - ν_{itn}^S),$$

where $γ_{itn}^W$ is an indicator that household $i$ watches show $tn$, $F_{ε_{itn}^S}(\cdot)$ denotes the cumulative distribution function (CDF) of $ε_{itn}^S$.

Computation of $Pr\left(γ_{itn}^W = 1 \mid γ_{itn}^S = 1\right)$ involves integrating out $\{v_{itn}^S\}_{n}^t$ and $ν_{it}^O$ in $Pr\left(γ_{itn}^W = 1 \mid γ_{itn}^S = 1, \{v_{itn}^S\}_{n}^t, ν_{it}^O\right)$. The sampling order implies $v_{itn}^S$ and $v_{it}^O$ in Equation (5) are truncated respectively below $\bar{u}_{itn}^S + ν_{itn}^S - \bar{u}_{itn'}^S$ and below $\bar{u}_{itn}^S + ν_{itn}^S - \bar{u}_{it}^O$.

---

12For reasons indicated in §3.2.2, the duration of a sampling event is assumed to be between 30 seconds and 3 minutes. Within this interval, the choice set is assumed to be (and normally is) constant.
If $u^S_{itn} < IV_{it, X_0}$, household $i$ samples another show if the inclusive value of remaining shows is higher than the value of the outside good. The choice set is now $\mathcal{N}_i = \mathcal{N}_i \setminus n$, and the probability of sampling show $t\tilde{n}$ ($\tilde{n} \in \mathcal{N}_i$) is:

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

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

### 4.2.2 Watching

Upon selecting show $tn$ to view, household $i$ 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. (2013), Nevskaya and Albuquerque (2013)). If the external shock is sufficiently negative, the household terminates the show.\textsuperscript{13}

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

Under the assumption that these external shocks arrive via a homogeneous Poisson process with rate $\lambda_{itn}$ for household $i$ and show $tn$, the probability of household $i$ exiting the show $tn$ at time $t'$, $q_{iitn'}$, is given by the probability that the flow utility upon receiving a new viewing shock falls below the alternative options.\textsuperscript{14}

$$q_{iitn'} \mid \left\{ v^S_{itn}, v^O_{it} \right\} = Pr \left( \bar{u}^S_{itn} + v^S_{itn} + e^S_{it'n} < \max \left\{ \bar{u}^O_{it} + v^O_{it}, ln \left( \sum_{n' \in \mathcal{N}_i'} \exp (\bar{u}^S_{it'n'} + v^S_{it'n'}) \right) \right\} \right) \hspace{1cm} (7)$$

\textsuperscript{13}This characterization of the show exiting decision is in essence similar to the First-Hitting-Time (FHT) models, see Lee and Whitmore (2006) for a comprehensive review of the literature.

\textsuperscript{14}The homogeneous Poisson process implies that within a show, external shocks arrive at a constant rate for the household.
\[
F_{\epsilon_{it}^n} \left( \max \left\{ \bar{u}_{it}^O + \nu_{it}^O, \ln \left( \sum_{n' \in \mathcal{N}_{it}'} \exp \left( \bar{u}_{it'n'}^S + \nu_{it'n'}^S \right) \right) \right\} - \bar{u}_{itn}^S - \nu_{itn}^S \right).
\]

Computation of \( q_{itn'} \) involves integrating out \( \{ \nu_{itn'}^S \}_n \) and \( \nu_{itn'}^O \) in \( q_{itn'} | \{ \nu_{itn'}^S \}_n, \nu_{itn'}^O \). \( q_{itn'} \) is not necessarily fixed through the duration of show \( tn \), and can alter when available shows on alternative networks \( (N_{it'}) \) change. For instance, when a new show \( t'n' \) starts on network \( n' \), \( \bar{u}_{it'n'}^S + \nu_{it'n'}^S \) changes and \( q_{itn'} \) would change accordingly. Hence, \( q_{itn'} \) is piecewise constant and changes whenever shows on alternative networks \( (N_{it'}) \) change. Note that in between shocks, \( q_{itn'} \) remains fixed.

Using an approach similar to Arcidiacono et al. (2013) and Nevskaya and Albuquerque (2013), Appendix A shows for \( q_{itn'} \) that is piecewise constant with segment 1, \( \ldots, M \), the CDF of the viewing length \( l_{itn}^* \) is:

\[
F_{l_{itn}^*} (\bar{t}) \equiv Pr \{ l_{itn}^* \leq \bar{t} \} = 1 - e^{-\lambda_{itn} \sum_{m=1}^{M} l_{itnm} q_m},
\]

where \( q_m \) is the exiting probability in segment \( m \), and \( l_{itnm} \) is the length of segment \( m \) up to time \( \bar{t} \), \( \sum_{m=1}^{M} l_{itnm} = \bar{t} \).

The probability density function of \( l_{itn}^* \) is therefore:

\[
f_{l_{itn}^*} (\bar{t}) = \lambda_{itn} \bar{m} q_{itnm} e^{-\lambda_{itn} \sum_{m=1}^{M} l_{itnm} q_m},
\]

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

As it is not possible to watch the show past its end,

\[
l_{itn}^W = \min (l_{itn}^*, L_{itn}),
\]

where \( l_{itn}^W \) is the time household \( i \) spends watching show \( tn \), and \( L_{itn} \) is the remaining length of show \( tn \) when sampled.

### 4.3 TV Show Recording

As DVRs are typically filled to capacity, the decision to record a new show is accompanied by the deletion of an older recorded show, thus revealing information about show preferences.
The TiVo DVR used by the panelists records one show at a time. Therefore, a newly recorded show $tn$ is assumed to have i) higher expected flow utility than the show that is replaced ($u_{itn}^S > u_{itd}^S$); and ii) higher expected flow utility than all shows that air at time $t$ but are not recorded ($u_{itn}^S > u_{itn'}^S, \forall n' \neq n$). As the show deleted is nearly always automatically selected by the DVR, we refrain from using the deletion choice decision.

Based on the flow utility specified in Equation (1), conditions i) and ii) imply the probability that household $i$ records show $tn$ is given by:

$$
Pr (y_{itn}^R = 1) = \frac{\exp (\tilde{u}_{itn}^S)}{\exp (\tilde{u}_{itd}^S) + \sum_{n'} \exp (\tilde{u}_{itn'}^S)},
$$

(11)

where $y_{itn}^R$ is an indicator that show $tn$ is recorded by household $i$.

### 4.4 TV Advertising Viewing

TV advertising avoidance differs between live and recorded viewing. In live viewing, advertising avoidance involves switching away from the advertisement (zapping). Hence the viewer’s alternative set includes other shows. In contrast, almost all advertising avoidance when views are recorded involves forwarding (zipping). Hence, the viewers’ alternative set includes the opportunity cost of time. Below, we formalize these points.

#### 4.4.1 Zapping

In the viewing model introduced above, the household can choose to avoid advertisements in a live show by channel switching (zapping). Similar to §4.2.2 where households can make channel switching decisions during program content upon receiving an external shock, households make zapping decisions during each advertisement. Hence, this component of the viewing model is analogous to the watching model, except we observe the specific point at which the utility changes. The zapping decision depends on the relative attractiveness of the advertisement as compared with 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:

$$
Pr \left( y_{itn}^{AL} = 0 \mid \left\{ v_{itn}^S \right\}_n, v_{it}^O \right)
$$
\[ Pr \left( \bar{u}^A_{itn} + \epsilon^A_{itn} < \max \left\{ \bar{u}^O_{it} + \nu^O_{it}, \ln \left( \sum_{n' \in N_{it}} \exp \left( \bar{u}^S_{itn'} + \nu^S_{itn'} \right) \right) \right\} - c_i \right) = F_{\epsilon^A_{itn}} \left( \max \left\{ \bar{u}^O_{it} + \nu^O_{it}, \ln \left( \sum_{n' \in N_{it}} \exp \left( \bar{u}^S_{itn'} + \nu^S_{itn'} \right) \right) \right\} - c_i - \bar{u}^A_{itn} \right), \]  

(12)

where \( y^A_{itn} \) is an indicator that the live advertisement \( tn \) is viewed (not zapped) by household \( i \), \( c_i \) is the zapping cost faced by household \( i \).\(^{15}\) Computation of \( Pr (y^A_{itn} = 0) \) involves integrating out \( \{\nu^S_{itn}\}_n \) and \( \nu^O_{it} \) in \( Pr (y^A_{itn} = 0 | \{\nu^S_{itn}\}_n, \nu^O_{it}) \).

### 4.4.2 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 household watches it. The probability that household \( i \) zips a recorded (R) advertisement \( tn \) is:

\[ Pr \left( y^A_{itn} = 0 \right) = \frac{\exp \left( \bar{u}^O_{it} \right)}{\exp \left( \bar{u}^O_{it} \right) + \exp \left( \bar{u}^A_{itn} \right)}, \]  

(13)

where \( y^A_{itn,R} \) is an indicator that the recorded advertisement \( tn \) is viewed (not zipped) by household \( i \).\(^{16}\)

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 and identification in the next section.

\(^{15}\)We consider two zapping costs - zapping during the show and zapping during the advertisement. Zapping costs can reflect the psychological cost associated with channel switching. For shows, this zapping cost cannot be separately identified from the arrival rate of external shocks. A household that switches channel 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 subsection.

\(^{16}\)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 in essence measures the relative cost of zapping versus zipping, and is identified from the difference in zipping and zapping probabilities.
5 Estimation and Identification

All parameters are household-specific as indicated by subscript $i$. We perform estimation household-by-household, facilitated by the availability of panel data of relatively long cross-section and duration for each household.\footnote{On average, there are 1,313 sampling occasions and 1,993 advertising viewing occasions per household.}

5.1 Estimation

The viewing model is estimated by simulated maximum likelihood. The likelihood is derived in Appendix B.

As virtually every household watches only a handful of available networks, we construct household-specific consideration sets based on viewing history. For each household, 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: i) live shows that are available on networks within the consideration set; and ii) shows stored on DVR that are recorded either manually or through a season pass.

To limit the size of parameters governing show and advertising flow utilities ($\beta^S_i$ and $\beta^A_i$), we only estimate flow utility parameters associated with the 6 most popular genres (drama, comedy, reality TV, talk shows, news, and sports, together accounting for 68% of viewing time) and the 6 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 advertisement sample. In an initial effort to capture the effects of product category on advertising preference, we classify the 574 product categories into four general categories: consumer packaged goods (CPG), service, drug, and other goods.

The arrival rate of external shocks, $\lambda_{itn}$, is parameterized as a function of genre:

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

where $g_{tn}$ is a row vector on genre, the $j$th element being an indicator variable of whether show $tn$ is of the $j$th genre.
5.2 Identification

Parameters that govern the flow utility of the show ($\beta_i^S$) and the outside good ($\beta_i^O$) are jointly identified by observed sampling, watching, and recording decisions, all revealing show preferences and time preferences. Similarly, parameters related to the flow utility from advertisements ($\beta_i^A$) are identified by variation in zipping and zapping decisions on advertisements.

Because the flow utility of the outside good varies by day, and the available programs usually also vary by day, one difficulty is to separate the value of the outside good from show quality in a day. If household $i$ does not watch TV on a day, then there are two possibilities. For the first, all shows available are associated with low flow utility (i.e., small $u_{itn}^S$ for all $t, n$). For the second, the opportunity cost of time is too high (i.e., large $u_{it}^O$). The recording behavior helps us disentangle these two factors. If no show is watched on a day but one or more shows are recorded, then it follows that the main reason for no viewing is high opportunity cost of time. If, on the other hand, none of the shows is watched or recorded, then the main reason for no viewing is low expected show quality in the day.

The parameters that determine the arrival rates of external shocks ($\rho_i$), and in turn, $\lambda_{itn}$, are identified by time spent on different types of shows. For instance, if a household switches more often in news than in dramas, then the household has a higher $\lambda_{itn}$ if show $tn$ is a news show than if it is a drama show, all else equal.

One concern is that people may switch less in shows that are more preferred, which leads to the question of whether the flow utility ($\beta_i^S$) and the shock arrival rate ($\lambda_{itn}$) can be separately identified. While sampling (show choice) decisions depend only on flow utility, viewing length depends on both shock arrival rate and flow utility. For instance, two shows with equal sampling probabilities can be watched for different lengths. All else equal, the show with a longer viewing length is associated with a lower shock arrival rate.

Finally, the zapping cost ($c_i$) is identified from the difference in zipping and zapping probabilities.

To assess whether the proposed estimation approach can recover known parameters, we develop a simulated dataset and implement the proposed estimation approach (Appendix B). The results show that the estimation approach works well in recovering known parameters.
6 Estimation Results

This section reports the estimation results of the viewing model. We use data from July 2005 to June 2006 for estimation, and reserve July 2006 for the policy experiments.

The second column of Table 3 reports the median (across households) of parameter estimates in flow utilities of shows, as well as the percentages of households with significantly positive and negative estimates (5% level) respectively. This column indicates the average viewer prefers shows that are short, familiar, live, and recently started (i.e., smaller viewing offset). The third column of this table reports the median (across households) of parameter estimates in flow utilities of advertisements, as well as the percentages of households with significantly positive and negative estimates (5% level) respectively. Results indicate viewers forward blocks of advertisements successively, as demonstrated in the positive coefficient regarding whether the preceding commercial is viewed. The first advertisement in a commercial break is less likely to be forwarded, presumably because it takes viewers some time to initiate a forwarding action. Once a forward starts, however, it tends to continue because the avoidance cost is sunk.
Table 3: Flow Utility Parameter Estimates for Shows and Advertisements

<table>
<thead>
<tr>
<th>Variable</th>
<th>Show Median Est (% Positive, % Negative, 5% level)</th>
<th>Advertisement Median Est (% Positive, % Negative, 5% level)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length</td>
<td>-0.21 (1.3%, 75.8%)</td>
<td></td>
</tr>
<tr>
<td>Number of episodes sampled in previous week</td>
<td>0.28 (83.9%, 0.6%)</td>
<td></td>
</tr>
<tr>
<td>Live</td>
<td>9.96 (86.5%, 0.0%)</td>
<td></td>
</tr>
<tr>
<td>Viewing offset</td>
<td>-4.78 (0.1%, 88.4%)</td>
<td></td>
</tr>
<tr>
<td>Watch the preceding ad</td>
<td></td>
<td>5.21 (74.0%, 2.3%)</td>
</tr>
<tr>
<td>First ad break</td>
<td>-0.11 (10.2%, 20.7%)</td>
<td></td>
</tr>
<tr>
<td>Last ad break</td>
<td>0.05 (17.9%, 15.4%)</td>
<td></td>
</tr>
<tr>
<td>First slot in a break</td>
<td>1.50 (63.9%, 2.5%)</td>
<td></td>
</tr>
<tr>
<td>Last slot in a break</td>
<td>-0.14 (0.06%, 0.12%)</td>
<td></td>
</tr>
<tr>
<td>Genre: Drama</td>
<td>-0.22 (14.5%, 43.2%)</td>
<td>-0.79 (14.6%, 31.5%)</td>
</tr>
<tr>
<td>Genre: Comedy</td>
<td>-0.48 (9.2%, 57.2%)</td>
<td>-0.33 (15.4%, 21.1%)</td>
</tr>
<tr>
<td>Genre: Reality TV</td>
<td>-0.69 (8.4%, 55.0%)</td>
<td>-0.88 (13.7%, 30.5%)</td>
</tr>
<tr>
<td>Genre: Talk shows</td>
<td>0.11 (34.9%, 25.3%)</td>
<td>-0.28 (13.3%, 18.6%)</td>
</tr>
<tr>
<td>Genre: News</td>
<td>0.06 (29.0%, 23.9%)</td>
<td>-0.67 (11.8%, 20.9%)</td>
</tr>
<tr>
<td>Genre: Sports</td>
<td>-0.46 (11.8%, 39.5%)</td>
<td>-0.47 (11.5%, 21.8%)</td>
</tr>
<tr>
<td>Network: ABC</td>
<td>0.73 (48.6%, 13.7%)</td>
<td>-0.20 (17.5%, 22.3%)</td>
</tr>
<tr>
<td>Network: CBS</td>
<td>0.54 (42.6%, 20.0%)</td>
<td>-0.29 (15.8%, 21.0%)</td>
</tr>
<tr>
<td>Network: NBC</td>
<td>0.80 (52.0%, 17.5%)</td>
<td>-0.35 (15.2%, 24.7%)</td>
</tr>
<tr>
<td>Network: FOX</td>
<td>0.14 (27.2%, 21.8%)</td>
<td>-0.25 (13.6%, 19.3%)</td>
</tr>
<tr>
<td>Network: USA</td>
<td>-0.26 (8.5%, 14.3%)</td>
<td>-0.16 (8.6%, 8.5%)</td>
</tr>
<tr>
<td>Network: Comedy Central</td>
<td>-0.20 (9.0%, 11.2%)</td>
<td>0.78 (6.1%, 4.1%)</td>
</tr>
<tr>
<td>Product category: CPG</td>
<td></td>
<td>-0.00 (0.09%, 0.08%)</td>
</tr>
<tr>
<td>Product category: service</td>
<td></td>
<td>0.03 (0.09%, 0.06%)</td>
</tr>
<tr>
<td>Product category: drug</td>
<td></td>
<td>-0.06 (0.04%, 0.07%)</td>
</tr>
</tbody>
</table>

There also exists extensive heterogeneity in zapping cost across households, meaning some households do not avoid live advertisements (and presumably are better targets than those that do). The mean, median and the standard deviation of the zapping cost are respectively 2.5, 1.0 and 4.1, and the 2.5% and 97.5% quantiles are respectively 0.01 and 2.8.

Table 4 reports the mean estimates of shock arrival rates by genre. The average viewer is more likely to switch channel 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 (2000)’s finding that viewing persistence is higher for dramas and lower for news and sports.
Overall, there exists considerable heterogeneity across households 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.\textsuperscript{18}

7 Policy Experiments: Advertising Targeting

This section conducts advertising targeting policy experiments. We first link advertising viewing with sales response in §7.1. Predicated on viewing behavior and advertising response, we discuss various targeting strategies in §7.2.

7.1 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 sufficient cross-sectional and temporal variation in both purchase and advertising. Within these categories, we consider 22 leading brands that have non-negligible market share and advertised during the sample period.

We apply a “model-free” approach to measure advertising effects.\textsuperscript{19} We start with homogeneous advertising effects, then allow advertising effects to vary across consumer segments.

Homogeneous Advertising Effects  A key issue with the measurement of advertising response is that advertising effects tend to be small relative to those of pricing and other potentially confounding factors

\textsuperscript{18}We performed several model validity checks using the hold-out sample of July 2006. Results indicate the proposed model outperforms a null model with equal flow utilities across shows in predicting sampling choices and watching choices. The proposed model also outperforms a null model with constant exiting rate throughout the show in predicting length of viewing. Detailed results are available in Online Appendix C.

\textsuperscript{19}We also estimate a latent class model of advertising response. It is worth noting that we find 36\% of all campaigns have a positive and significant advertising effect at p=0.4. This percentage is exactly identical to that reported for mature consumer packaged goods in the meta-analysis of cable advertising experiments in Lodish et al. (1995).
Moreover, the attribution of a sale to a particular advertisement is challenging and often involves various assumptions about decay (e.g., Clark et al. (2009)). Because of these problems, researchers typically measure advertising effects by using multivariate models to control for other covariates and past advertising. However, this requires a number of assumptions regarding functional form, model specification, and error distributions.

As an alternative, we seek to hold the in-store causal environment constant by considering two same-store visits within the same week, where a household did not purchase on the first trip. As store causals are held constant, this removes a major source of variation in purchase behavior. Also, as the contrast is within household, comparing second and first trips removes unobserved individual effects, including advertising prior to the first visit. Two key factors that remain are inventory and advertising between visits. With regard to inventory, draw down is limited in one week meaning that inventory levels should be similar on the second visit. Hence, our strategy is to compare the difference in second-visit demand when advertising appears between weekly visits and when advertising does not. As long as advertising is not highly correlated with inventory draw down between visits, this should yield a model free, non-parametric estimate of advertising effects.

Using this approach, we compute each brand’s second-visit purchase log odds ratio with and without advertising exposures received between the two visits (Figure 9). Under the assumption of no advertising effect, the log odds ratio should be centered around zero. As the distribution is instead skewed to the right of zero, it suggests a positive advertising effect for most brands. Reflective of this overall positive advertising effect, a one-sample Kolmogorov-Smirnov test indicates the brand-level log odds ratio is not normally distributed with mean zero (p<0.016). Overall, advertising effects are positive, but small.

---

20 A category purchase in the first store visit makes a category purchase in the second store visit very unlikely. If three or more same-store visits are observed within the same week, each visit is paired with its preceding visit.

21 As a robustness check, we address the concern that the duration between the two visits might increase both advertising exposure and purchase likelihood independently of advertising. When the second visit is further apart from the first visit, the household is more likely to be exposed to advertising due to more TV viewing, and might also be more likely to purchase in the second visit due to inventory reduction. Accordingly, we compute purchase log odds ratios by brand and the duration between store visits (1-2 days, 3-4 days, and 5-6 days). We find that the distribution within each duration group is skewed to the right of zero. As the durations are held fixed in each group, variation in duration cannot be the cause of the advertising lift.
Heterogeneous Advertising Effects  Because effective targeting relies upon heterogeneity in advertising response, we seek to categorize consumer responsiveness heterogeneity using a discrete mixture approach owing to its parsimony and flexibility. All else equal, consumers who are more responsive to advertising make better targets. For household $i$ in segment $k$ ($k = 1, \ldots, K$), the second-visit probability of purchasing brand $j$ is specified as:

$$
Pr\left(y_{ijm}^P = 1 \mid i \in k\right) = \frac{\exp \left( \theta_k A_{ijm} + \alpha_{jk} \right)}{1 + \exp \left( \theta_k A_{ijm} + \alpha_{jk} \right)},
$$

where $m$ denotes shopping trip, $y_{ijm}^P$ is an indicator variable of whether household $i$ purchases brand $j$ in shopping trip $m$. $A_{ijm}$ is an indicator variable of whether household $i$ has viewed brand $j$’s advertisement since the preceding trip (in the same store and week). $\alpha_{jk}$ is a brand intercept for brand $j$ and segment $k$.

The prior probability that household $i$ is in ad-response segment $k$ is specified as:

$$
Pr(i \in k) = \frac{\exp \left( \eta_k \right)}{\sum_{k'=1}^{K} \exp \left( \eta_{k'} \right)},
$$

where $\eta_1$ is normalized to zero for identification. Household $i$’s posterior probability of belonging to segment $k$ can therefore be obtained by Bayes’ rule:
The likelihood function associated with household $i$’s purchase decisions is:

$$L_i = \sum_k Pr(i \in k) \prod_m \prod_j (Pr(y_{ijm} = 1 \mid i \in k)) y_{ijm}$$

and the overall log likelihood is:

$$lnL = \sum_i lnL_i.$$ (19)

Because different product categories are associated with different purchase frequencies, we estimate this model separately for each product category, and the number of segments is determined based on BIC.

For three product categories (children’s yogurt, children’s cereal, and toothpaste), a single segment is identified. For two product categories (bathroom tissue and sports drink), two segments are identified. For the other two product categories (regular cola and diet cola), three segments are identified. Table 5 reports the estimated advertising effect ($\theta_k$) by product category and consumer segment. The estimates will be used in implementing household-level targeting strategies.

<table>
<thead>
<tr>
<th>Product Category</th>
<th>Segment 1</th>
<th>Segment 2</th>
<th>Segment 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Children’s yogurt</td>
<td>-0.03 (0.52)</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Children’s cereal</td>
<td>0.23 (0.53)</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Regular cola</td>
<td>0.51 (0.38)</td>
<td>0.20 (0.44)</td>
<td>-0.86 (0.58)</td>
</tr>
<tr>
<td>Diet cola</td>
<td>0.02 (0.28)</td>
<td>-0.68 (0.55)</td>
<td>-0.71 (0.69)</td>
</tr>
<tr>
<td>Sports drink</td>
<td>0.89 (0.33)</td>
<td>-0.12 (0.73)</td>
<td>NA</td>
</tr>
<tr>
<td>Toothpaste</td>
<td>-0.18 (0.31)</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Bathroom tissue</td>
<td>0.27 (0.27)</td>
<td>-0.03 (0.32)</td>
<td>NA</td>
</tr>
</tbody>
</table>

Note: Standard errors in parentheses.

Overall, the short-term effects of advertising on sales appear to be small, and there is modest heterogeneity in advertising response.
7.2 Targeting Approaches

Currently, national TV networks sell advertising inventory in advance by show. In this “upfront” 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 commercials 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 viewers in a show as well as the demographic mix.

Digital distribution offers two further advances upon the upfront model. First, this technology allows advertisers to buy users instead of shows. As the viewing and advertising avoidance models developed in §4 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 in upfront markets. Second, advertisements can be inserted real time, analogous to current practices in Internet advertising. In this case, viewing is known at the time of advertisement insertion and purchase, but the likelihood of avoiding a subsequent advertisement must be forecasted.

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

Table 6 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. The entry in each cell refers to the specific sections in which we detail the alternative policy experiments and results.

<table>
<thead>
<tr>
<th>Table 6: Targeting Scenarios</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Real-time Buy</td>
</tr>
<tr>
<td>Advance Buy</td>
</tr>
</tbody>
</table>
We explore these policies using the leading brand of bathroom tissue, Charmin over the hold-out period of July 2006.22

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 competitors might adjust their own advertising schedules in response. While a fruitful area for additional research, we believe this analysis extends beyond the scope of this paper, and interpret our results in light of this caveat.

7.2.1 Cost-Based Real-time Buy

We first consider the potential for Charmin to lower its cost of these views (views differ from exposures inasmuch exposures need not be viewed). To do this, we first compute each targeted household \( i \)'s observed average per-exposure advertising cost, \( c_i \) (by averaging the observed advertising costs across all exposures to that household in the holdout data, \( c_{itn} \)). Next, we convert the cost per exposure into a predicted cost per view, as those who miss exposures are not efficient targets. We compute the predicted cost per exposure by dividing the observed exposure cost by the predicted viewership. For example, if the likelihood of viewing a particular exposure is 1/2, then the predicted cost per view is twice the cost per exposure. Denote the obtained average cost per view as \( c^\star_i \), and the predicted cost for a particular view on network \( n \) at time \( t \) as \( c^\star_{itn} \). Intuitively, if \( c^\star_{itn} < c^\star_i \), then a more efficient allocation of expenditures is feasible. This more efficient allocation would happen if the cost for a particular exposure is low relative to the average exposure, or if the advertisement viewership likelihood for that exposure is relatively high. Therefore, a parsimonious rule for the advertiser would be to set a level \( K \), conditioned on the information available at time \( t \), such that the advertiser purchases advertising slot \( tn \) if \( c^\star_{itn} < Kc^\star_i \). Lower values of \( K \)

22We focus upon bathroom tissue as it has broad penetration and little fluctuation in demand. Hence, stockpiling effects and forward purchase are less common. Moreover, there exists moderate heterogeneity in household advertising response in the bathroom tissue category (Table 5) which implies potential gains to targeting. Within this category, we focus upon Charmin as it has the largest number of advertising exposures (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 hold-out period of July 2006, the 834 sample households viewed 690 Charmin’s advertisements at a total cost of $8.77.
imply a rule that leads to higher levels of purchase efficiency, but at the cost of reach (there will only be a limited number of exposures with low costs and high viewing likelihoods). We implement this purchase rule viewer-by-viewer; starting at the beginning of the month such that Charmin buys all exposures until a) the period has ended, or: b) both advertising exposures and expenditures under the new rule exceed those observed under the current schedule.

Figure 10 portrays the total advertising views and costs (across households) under different values of $K$. In the figure, we note that total advertising costs decrease with $K$. This is primarily because a lower value for $K$ means that Charmin is buying advertisements that have lower costs per view. As the maximum cost per view decreases, the available advertising inventory meeting this criteria shrinks, and the total campaign expenditures accordingly decrease. 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 costs 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 if there is sufficient inventory because one can buy more views for a fixed budget. This implies increased views from more efficient buying. How these two opposing forces tradeoff is a function of the simple budget rule listed above (spend until time runs out, or both the reach and budget exceed that of the original schedule). If the reach criteria binds first, and the end of period criteria binds second, then it is possible for a) exposures to increase (because spending on exposures continues if the time or budget has not bound) and b) total budgets to decrease (because the end of the period arises before the total budget is spent).

When $K$ is set below 0.6 (implying high cost efficiency), advertising views surpass those under the current schedule even though overall campaign costs decrease. For example, when $K = 0.55$, Charmin’s exposures are increased by 47% while its costs are decreased by 12%. Charmin buys fewer advertisements, but the advertisements cost less and are more likely to be seen. Thus it is possible to lower costs and increase exposures. Of additional interest is the degree to which the real-time buy cost efficiency gains accrue from placing advertisements in shows that are currently being watched vs. the gains arise

\[23\] A generalization of the aforementioned parsimonious decision rule for advertising buys is to soften either the minimum advertising exposures or the maximum budget constraints. This exercise yields similar insights - that is, with $0.5 \leq K \leq 0.7$, it is possible to decrease costs and increase views on the same order of magnitude. Hence, as a general rule of thumb, setting $K$ to low levels (0.5 to 0.6) appears to enhance advertising purchase efficiency.
from targeting advertisements to those who are less likely to avoid them. To answer this question we replace the model’s forecasted advertising avoidance (denoted Model A) with the average advertising avoidance rate across viewers in the data (denoted Model B); we then use Model B to re-impute advertising costs and exposures under the simple buying rule. Again setting $K = 0.55$, our first finding is that it is no longer the case that exposures increase and costs decrease under Model B; instead, both increase. Second, the average costs per view in the observed data is 1.27 cents, under Model A is 0.76 cents, and under Model B is 0.85 cents. Thus, show placement alone yields a 33% improvement in cost efficiency over the observed schedule, but coupled with the advertisement viewing model there is a 40% improvement (in other words, roughly 1/5 of the improvement in efficiency is due to reducing advertising avoidance and the other 4/5 to reducing show avoidance).

Figure 10: Number of Advertising Views and Costs Under Different Purchase Thresholds ($K$). The current observed cost is $8.77 with 690 views.

7.2.2 Cost-Based Advance Buy

In this targeting scenario, the advertiser purchases advertising in advance for a given duration, minimizing the total cost of expected exposures for targeted households during live viewing. In this scenario, advertisers are unable to condition on current viewing meaning that they need to predict show viewing as
well as advertising viewing. Hence, we assume advertisers seek to buy a single advertising slot sometime
within the show rather than seeking to buy during a specific pod.\textsuperscript{24}

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

$$\text{Min}_{\{x_{itn}\}} \sum_{t} \sum_{n} c_{tn} x_{itn}$$

\text{s.t.} \quad x_{itn} \in \{0, 1\}, \forall t, n, \quad \text{(20)}

$$E\left(\sum_{i} \sum_{n} x_{itn} r_{itn}\right) \geq E\left(\sum_{i} \sum_{n} x_{itn}^* r_{itn}\right),$$

\text{(22)}

where $x_{itn}$ denotes household-show selection under the optimal schedule. $x_{itn} = 1$ if show $tn$ is selected
for household $i$, $x_{itn} = 0$ otherwise. $x_{itn}^*$ denotes household-show selection under the current schedule.
$x_{itn}^* = 1$ if show $tn$ is selected for household $i$, $x_{itn}^* = 0$ otherwise. $c_{tn}$ denotes the per-exposure advertising
price associated with show $tn$. $r_{itn}$ 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 same viewers
as it currently targets.

The expectations in Equation (22) are taken over the distribution of $r_{itn}$, which is associated with
the uncertainty in advertising viewing estimates. $r_{itn}$ 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,

$$r_{itn} = \Pr(y_{itn}^S = 1) \Pr(y_{itn}^W = 1 | y_{itn}^S = 1) \times \left(\int_{t'} \Pr(l_{itn}^W \geq t') \Pr(y_{itn}^A = 1) f(t') \, dt'\right),$$

\text{(23)}

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 timing of advertisements 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 $l_{itn}^W$ exceeds $t'$. $\Pr(y_{itn}^A = 1)$ denotes the probability that this advertisement
will not be skipped.

\textsuperscript{24}We assume no network guarantee on minimal views to advertisers as is often common in practice.
Because $E \left( \sum_n x_{itn} r_{itn} \right) = \sum_t \sum_n x_{tn} E \left( r_{itn} \right)$, we first compute $E \left( r_{itn} \right)$ by simulation,\(^{25}\) and then use it to solve the optimization problem for each household. The average per-view advertising price reduces from 9.4 cents to 1.6 cents. The overall cost reduces from $8.77 to $1.48, a 83% reduction in expenses. Scaled to the 1.1 million TV households in the three DMAs where the data were collected, the cost reduction will be $9,615.

### 7.2.3 Profit-Based Real-time Buy

To ascertain the effect of advertising on profits, we compare the marginal effect of advertising to a given consumer in a given week to its cost. A profitable exposure is one wherein the marginal revenue exceeds the cost. The marginal cost of advertising is readily available in our data.

Incremental profits from advertising are computed using the model-free estimates from §7.1. Following Equation (15), the effect of viewing an advertisement for household $i$ on the purchase probability for brand $j$ is:

$$
\Delta_{ij} \mid i \in k = Pr \left( y_{ij}^P = 1 \mid i \in k, A_{ij} = 1 \right) - Pr \left( y_{ij}^P = 1 \mid i \in k, A_{ij} = 0 \right) 
= \frac{exp \left( \theta_k + \alpha_{jk} \right)}{1 + exp \left( \theta_k + \alpha_{jk} \right)} - \frac{exp \left( \alpha_{jk} \right)}{1 + exp \left( \alpha_{jk} \right)}.
$$

This lift is conditional on store visitation. To obtain an unconditional measure of advertising response, we use the household’s empirical distribution of weekly store visits to simulate shopping likelihoods under the assumption that an advertising view is not likely to affect shopping trips. The unconditional lift is given by the number of simulated trips times the conditional lift per trip.

The lift yields incremental revenues, not profits. As we do not observe unit margins, we solve for the profitable impressions under different unit dollar margins $m$.\(^{26}\) The expected lift of advertisement $tn$ is calculated as $Pr \left( y_{itn}^A = 1 \right) \Delta_{ij} N_{iw} m - c_{tn}$, where $Pr \left( y_{itn}^A = 1 \right)$ denotes the probability that the advertisement will not be zapped (if the show is live) or zipped (if the show is recorded), $N_{iw}$ denotes the number

\(^{25}\)Specifically, we draw 50 sets of household-specific parameter estimates, each leading to a prediction on households’ show choices and viewing length conditional on choice. $E \left( r_{itn} \right)$ is taken as the average of this predicted viewership across the 50 sets of parameter draws.

\(^{26}\)Bathroom tissue has gross profit margin of roughly 13%. In the data, the mean and median prices of Charmin are respectively $6.8 and $6.3, leading to a unit dollar margin of about $0.8-0.9.
of simulated trips in week $w$, and $c_{tn}$ is the cost of advertising slot $tn$. Advertisement $tn$ is purchased if its expected lift is positive.

Figure 11 depicts the simulated profits under the observed and proposed advertising schedules for these different margins. The simulated short-term advertising profit using the observed schedule is negative, with losses ranging from $6.2$ to $7.3$ across the 834 households in our sample. The proposed schedule produces a small, but positive simulated incremental profit (the gain ranges from 15 cents to 95 cents). The key rational for this increase is that most advertisements have negative short-term profitability. Hence, the proposed advertising schedule considerably cuts overall spending from the observed $8.77$ down to $0.65 - 2.78$ (with $0.60 - 1.00$ margins respectively). In other words, as most advertising is unprofitable in the short-term, it gets cut.

We conclude by noting that advertisers often set campaigns for longer term branding goals, hence there may be additional rationales to advertise other than short-term profits. Nonetheless, the outlined approach highlights the most profitable exposures over the short run.

![Figure 11: Simulated Incremental Profit (for Sample Households, Real-time Buy) Under Different Unit Dollar Margins](image)

Upon sharing our results with TiVo and Nielsen/Catalina, these firms indicated similar findings.

---

27Upon sharing our results with TiVo and Nielsen/Catalina, these firms indicated similar findings.
7.2.4 Profit-Based Advance Buy

As noted in §7.2.3, the expected incremental profit from an advertisement is \( Pr(\gamma_{itn}^A = 1) \Delta_{ij}N_{iw}m - c_{tn} \). Summing over households and weeks and replacing \( Pr(\gamma_{itn}^A = 1) \) with \( r_{itn} \) defined in Equation (23), the optimization problem of allocating exposures is given by:

\[
\text{Max}_{\{x_{itn}\}_{i,t,n}} \sum_{i} \sum_{t} \sum_{n} E \left\{ \sum_{w} \sum_{u} x_{itn} r_{itn} \Delta_{ij} N_{iw} m - \sum_{w} \sum_{n} x_{itn} c_{tn} \right\} \\
\text{s.t. } x_{itn} \in \{0, 1\}, \forall i, t, n.
\]

(25)

(26)

where \( x_{itn} \) denotes household-show selection. \( x_{itn} = 1 \) if show \( tn \) is selected for household \( i \), \( x_{itn} = 0 \) otherwise.

The expectation in Equation (25) is taken over the distribution of \( r_{itn} \), \( \Delta_{ij} \) and \( N_{iw} \). Because \( E \{ \sum_{t \in w} \sum_{n} x_{itn} r_{itn} \Delta_{ij} N_{iw} m \} = m \sum_{t \in w} \sum_{n} x_{itn} E (r_{itn} \Delta_{ij} N_{iw}) \), we first numerically compute \( E (r_{itn} \Delta_{ij} N_{iw}) \), then solve the optimization problem under different unit dollar margins \( m \). Overall spending reduces from the observed $8.77 down to a range of $0.01 - $0.18 (with the corresponding $0.60 - $1.00 margins). The simulated incremental profit ranges from 0.3 cents to 4.5 cents. In essence, this schedule cuts advertising to nearly zero - as in expectation, most advertising is unprofitable because of low advertising elasticities. The budgets are not cut so severely in the case of real-time buying, because those instances wherein one is certain a user will be exposed can increase the chance the advertisement has a positive outcome.

7.2.5 Targeting Summary

The targeting results are summarized in Table 7.

<table>
<thead>
<tr>
<th></th>
<th>Reduce Costs</th>
<th>Increase Profits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real-time</td>
<td>Target exposures increase as much as 47% while costs decrease by 12% in one case</td>
<td>Spending cut from $8.77 to $0.65-$2.78; profit increases from $-6.2-$7.3 to $0.15-$0.95</td>
</tr>
<tr>
<td>Advance</td>
<td>Costs reduce 83% for same expected exposures</td>
<td>Negligible spending and profit</td>
</tr>
</tbody>
</table>

Results suggest there are large percentage-wise decreases in costs and increases in exposures possible as a result of improved targeting. Further, short-term profits can be increased by sharply reducing
advertising, as the low advertising elasticities imply the ROI (Return on Investment) from short-term advertising is often likely to be negative. We further note that the allocation of surplus among multiple advertisers and TV networks in the long run depends on competition on both the product market and the advertising market, as well as the advertising pricing and allocation mechanisms.

8 Conclusion

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

In this paper, we use a unique dataset that integrates several data sources (DVR data, purchase data, show data and advertising data) to develop and estimate models of households’ TV program sampling/consumption and advertising response, and then conduct counterfactual policy experiments to evaluate potential gains from targeting. The viewing model is predicated upon a series of observations about the tendency of viewers to sample shows before viewing them, and avoid advertisements that appear within those shows. One key insight is that variation in advertising avoidance is greater across viewers than genres, times and other factors, suggesting that household level targeting can improve targeting efficiency.

Accordingly, we consider several household-level targeting scenarios by manipulating: 1) whether the objective function is to minimize costs for a given set of exposures or to maximize incremental profit from advertising; and 2) whether the advertising purchase is made in advance or in real time. Results indicate micro-targeting can lower advertising costs and/or raise incremental revenue.
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 advertisements placed within a program of dissimilar content are recalled significantly better than if 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 the targeting strategies.

Second, the targeting strategies we proposed do not take into account potential complementarities between the consumption of goods and advertisements (Becker and Murphy (1993)). According to a recent study by Tuchman et al. (2015), the quantity of the advertised product purchased recently can explain the advertising skip rates. They suggest that advertising efficacy would depend on the compliance of the targeted households, and the firm can target the subset of households whose purchases and welfare will likely change in response to the advertising campaign. Therefore, a fruitful extension to household-level targeting involves selection of households based on joint consumption of goods and advertisements. Another extension involves correlation between households’ preferences for show attributes and for products.

Third, the targeting strategies are designed based on an integrated channel structure, under which advertisers and TV networks share the same information and same objectives. Therefore, we are unable to model the pricing mechanism and the advertising allocation problem from the TV networks’ perspective. We do this because advertising allocation can be designed to align networks’ and advertisers’ incentives (e.g., Wilbur et al. (2013)). Still, another extension is to consider strategic interactions between advertisers and TV networks and how they affect gains available from micro-targeting.

Finally, future research can extend micro-targeting by taking into account competitive response and allowing for re-optimization 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


eMarketer. 2014. Advertisers blend digital and TV for well-rounded campaigns.


Online Technical Appendix

A Length of Viewing

Assume external shocks arrive via a homogeneous Poisson process (rate $\lambda$), and assume the exiting probability is piecewise constant with $M$ segments.

Further denote:

- $N_m(t)$: the total cumulative number of external shocks occurred during segment $m$ up to time $t$
- $N(t)$: the total cumulative number of external shocks up to time $t$, $N(t) = \sum_{m=1}^{M} N_m$
- $N_{exit}(t)$: the total cumulative number of external shocks that lead to exiting up to time $t$
- $p_m$: probability of keeping watching conditional on receiving an external shock during segment $m$
- $q_m(t)$: probability that a given shock is during segment $m$, which equals the share of segment $m$’s time up to time $t$, $t_m t$

Then we have:

\[
Pr \{ l \geq t \} = Pr \{ N_{exit}(t) = 0 \}
\]
\[
= \sum_{k=0}^{\infty} Pr \{ N(t) = k \} \sum_{i=1}^{k} \left[ Pr \{ N_1(t) = n_1, N_2(t) = n_2, \ldots, \} \right] \times \prod_{m=1}^{M} (p_m)^{n_m}
\]
\[
N_M(t) = n_M \left| \sum_{m=1}^{M} N_m(t) = k \right. \right. \times \prod_{m=1}^{M} (p_m)^{n_m}
\]
\[
= \sum_{k=0}^{\infty} e^{-\lambda t} (\lambda t)^k \frac{k!}{k!} \sum_{i=0}^{k} \frac{k!}{n_1! \ldots n_M!} q_1^{n_1}(t) \cdots q_M^{n_M}(t) \prod_{m=1}^{M} (p_m)^{n_m}
\]
\[
= \sum_{k=0}^{\infty} e^{-\lambda t} (\lambda t)^k \left( \sum_{m=1}^{M} q_m(t) p_m \right)^{n_m}
\]
\[
= \sum_{k=0}^{\infty} e^{-\lambda t} (\lambda t)^k \left( \sum_{m=1}^{M} q_m(t) p_m \right)^{n_m}
\]
\[
= \sum_{k=0}^{\infty} \left\{ \lambda t \left[ \sum_{m=1}^{M} q_m(t) p_m \right] \right\}^k
\]
\[
= e^{-\lambda t} \sum_{k=0}^{\infty} \left\{ \lambda t \left[ \sum_{m=1}^{M} q_m(t) p_m \right] \right\}^k
\]
\[
= e^{-\lambda t} e^{\lambda t \left[ \sum_{m=1}^{M} q_m(t) p_m \right]}
\]
\[ e^{-[1-\sum_{m=1}^{M} \frac{t_m}{\lambda_m} p_m]} \lambda t = e^{-\lambda \sum_{m=1}^{M} t_m (1-p_m)}. \]

## B Likelihood Function of the Viewing Model

We estimate the viewing model by simulated maximum likelihood approach. The sampling order implies that in Equations (5), (7), and (12) are truncated respectively below \( \bar{u}_{itn}^S + v_{itn}^S - \bar{u}_{itn}' \) and below \( \bar{u}_{itn}^S + v_{itn}^S - \bar{u}_{itn}^O \). In the estimation, we first simulate \( K = 100 \) sets (Chen and Yao (2015)) of \( \{v_{itn}^S\}_n, \{v_{itn}^O\} \), denoted as \( \{v_{itn}^{S1}\}_n, \ldots, \{v_{itn}^{SK}\}_n, \{v_{itn}^{O1}\}, \ldots, v_{itn}^{OK} \). We then derive the simulated likelihood associated with sampling, watching, recording and advertising viewing decisions, and aggregate them to obtain the overall likelihood function.

### Sampling

Each sampled show contributes to the likelihood with:

\[ Pr\left( y_{itn}^S = 1 \right) = \frac{\exp\left( X_{itn}^S \beta_i^S \right)}{\exp\left( X_{itn}^O \beta_i^O \right) + \sum_{n' \in \mathcal{N}_O} \exp\left( X_{itn}' \beta_i^O \right)}. \]

### Watching

Each sampled show that is watched contributes to the simulated likelihood with:

\[ Pr\left( y_{itn}^W = 1 \mid y_{itn}^S = 1 \right) = Pr\left( u_{itn}^S \geq IV_{itn} \right) = \frac{1}{K} \sum_{k=1}^{K} \left[ 1 - F_{\alpha_{itn}} \left( \max \left\{ u_{itn}^O + v_{itn}^{Ok}, \ln \left( \sum_{n' \in \mathcal{N}_O} \exp\left( u_{itn}' + v_{itn}' \right) \right) \right\} - \bar{u}_{itn}^S - v_{itn}^O \right) \right], \]

and each sampled show that is not watched contributes to the likelihood with:

\[ Pr\left( y_{itn}^W = 0 \mid y_{itn}^S = 1 \right) = 1 - Pr\left( y_{itn}^W = 1 \mid y_{itn}^S = 1 \right) \]

### Switching

Each show that is watched until the end (\( l_{itn}^W = L_{itn} \)) contributes to the likelihood with:
\[
Pr\left(l_{itn}^W \mid l_{itn}^W = L_{itn}\right) = Pr\left\{l_{itn}^* > L_{itn}\right\} \\
= \frac{1}{K} \sum_{k=1}^{K} e^{-\lambda_{itn} \sum_{m=1}^{M} q_{itn}^m \left(\nu_{itn}^S \right) } e^{-\lambda_{itn} \sum_{m=1}^{M} q_{itn}^m \left(\nu_{itn}^O \right) },
\]

where:
\[
q_{itn}^m \mid \left\{\nu_{itn}^S \right\}_{n'}, \nu_{itn}^O = F_{\nu_{itn}^S} \left(\max \left\{\bar{u}_{itn} + \nu_{itn}^O, \ln \left(\sum_{n' \in \mathcal{N}_{itn}'} \exp \left(\bar{u}_{itn'} + \nu_{itn'}^S \right) \right) \right\} \right) - \bar{u}_{itn} - \nu_{itn}^O.
\]

Each show that is not watched until the end \(l_{itn}^W < L_{itn}\) contributes to the likelihood with:
\[
Pr\left(l_{itn}^W \mid l_{itn}^W < L_{itn}\right) = f(l_{itn}^W) \\
= \frac{1}{K} \sum_{k=1}^{K} \lambda_{itn} \sum_{m=1}^{M} q_{itn}^m \left(\nu_{itn}^S \right) e^{-\lambda_{itn} \sum_{m=1}^{M} q_{itn}^m \left(\nu_{itn}^O \right) },
\]

where \(\bar{m}\) is the segment that \(l_{itn}^W\) falls into.

**Recording**

Each recorded show contributes to the likelihood with:
\[
Pr\left(y_{itn}^R = 1\right) = \frac{\exp\left(X_{itin}^S \beta_i^S\right)}{\exp\left(X_{itin}^S \beta_i^S\right) + \sum_{n'} \exp\left(X_{itin}^S \beta_i^S\right)}.
\]

**Advertising zapping (live shows)**

Each non-zapped advertisement contributes to the likelihood with:
\[
Pr\left(y_{itin,L}^A = 1\right) = 1 - \frac{1}{K} \sum_{k=1}^{K} \left[ F_{\nu_{itin}^S} \left(\max \left\{\bar{u}_{itn} + \nu_{itn}^O, \ln \left(\sum_{n' \in \mathcal{N}_{itn}'} \exp \left(\bar{u}_{itn'} + \nu_{itn'}^S \right) \right) \right\} \right) \\
- c_i - \bar{u}_{itin} \right],
\]

and each zapped advertisement contributes to the likelihood with:
\[
Pr\left(y_{itin,L}^A = 0\right) = 1 - Pr\left(y_{itin,L}^A = 1\right).
\]
Advertising zipping (recorded shows)

Each non-zipped advertisement contributes to the likelihood with:

$$Pr\left(y_{itn,R}^A = 1 \right) = \frac{\exp \left( X_{itn}^A \beta_i^A \right)}{\exp \left( X_{itn}^A \beta_i^A \right) + \exp \left( X_{itn}^O \beta_i^O \right)},$$

and each zipped advertisement contributes to the likelihood with:

$$Pr\left(y_{itn,R}^A = 0 \right) = 1 - Pr\left(y_{itn,R}^A = 1 \right).$$

Due to the high computational cost associated with simulation of \( \nu_{itn}^S \) and \( \nu_{it}^O \), we take two steps in the estimation. In the first step, we estimate \( \beta_i^S, \beta_i^O \) and \( \beta_i^A \) using likelihoods associated with sampling, recording and advertising zipping decisions. These likelihoods do not rely on \( \nu_{itn}^S \) and \( \nu_{it}^O \). In the second step, we estimate \( \rho_i (\lambda_i) \) and \( c_i \), taken estimates \( \hat{\beta}_i^S, \hat{\beta}_i^O \) and \( \hat{\beta}_i^A \) from the first step as given.

Further, to assess whether the proposed estimation approach can recover known parameters, we simulate a synthetic dataset and implement the proposed estimation approach on the simulated data. Table A.1 presents the results. The results show that the estimation approach works well in recovering known parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True Value</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_1^S )</td>
<td>15.0</td>
<td>14.58</td>
<td>0.70</td>
</tr>
<tr>
<td>( \beta_2^S )</td>
<td>-1.5</td>
<td>-1.44</td>
<td>0.13</td>
</tr>
<tr>
<td>( \beta^C )</td>
<td>6.0</td>
<td>5.86</td>
<td>0.33</td>
</tr>
<tr>
<td>( \beta_1^A )</td>
<td>9.0</td>
<td>8.76</td>
<td>0.50</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>0.5</td>
<td>0.48</td>
<td>0.09</td>
</tr>
<tr>
<td>( c )</td>
<td>20.0</td>
<td>19.76</td>
<td>0.80</td>
</tr>
</tbody>
</table>

C Model Validation

We perform several model validity checks for the viewing model using the hold-out sample of July 2006. The first examines the hit rates (by household) in show sampling and watching (conditional on sampling) predictions, which can be obtained using Equations (4) and (5). We consider hit rates under several alternative models for the flow utility of shows: including all covariates (the proposed model), excluding network covariates (model less network), excluding genre covariates (model less genre), and a null model with equal flow utilities across shows. Figure A.1 indicates for both sampling and watching
predictions, the hit rate under the proposed model first-order stochastically dominates the hit rate under the null model. Therefore, the proposed model performs better than the null model. In addition, networks are slightly more important than genres in predicting sampling and watching choices.

Figure A.1: Empirical CDF (Across Households) of the Hit Rate on Show Sampling (a) and Watching (b)

The second validity check concerns the viewing length (conditional on watching). Figure A.2 compares the mean absolute error (MAE) of viewing length predictions under the proposed model and under a null model where exiting rate is constant throughout the show. The proposed model also outperforms the null model as reflected in the first-order stochastic dominance relationship.

Figure A.2: Empirical CDF (Across Households) of MAE on Viewing Length