TV Viewing and Advertising Targeting

Yiting Deng and Carl F. Mela
Report Summary

Digital innovations such as digital video recorders and set-top boxes are both a blessing and a curse for advertisers. On the one hand, set-top boxes have greatly enhanced viewers’ TV-viewing experiences, leading to increased viewing consumption and the opportunity to target advertising at the TV-set level. On the other hand, set-top boxes enable viewers to fast-forward past, and thus avoid, advertisements.

In this report, Yiting Deng and Carl Mela take advantage of micro-level viewing data and micro-targeting capabilities inherent in set-top boxes to better understand viewer behavior and propose ways for advertisers to improve their targeting profitability.

They use a unique dataset that integrates several data sources (set-top box data, purchase data, show data, and advertising data) to develop and estimate models of households’ TV program sampling / consumption and advertising response.

Their viewing model is based on the following observations:

- Households watch prime-time TV almost every night (average = 85% of days) and for most of the evening (average = 88% of prime-time hours). Thus, any gains from targeting largely arise from what a viewer watches rather than whether they watch.

- Once a viewing session commences, viewers generally sample several shows prior to selecting one to view. Once a show is selected, viewers typically watch a show to its conclusion. Within a show, viewers’ advertising avoidance is more common when the show is recorded than when it is live.

- Person-specific factors explain most of the variation in advertising avoidance (20.4%), suggesting the potential for significant returns to household-level advertising targeting. Variations explained by genre, brand, time, and category are respectively 0.7%, 0.3%, 0.0%, and 0.0%, suggesting targeting on these will have smaller effects on advertising avoidance.

Deng and Mela consider several household-level targeting scenarios by manipulating: (1) whether the object is to minimize costs for a given set of views or to maximize incremental profit from advertising, and (2) whether the advertising purchase is made in advance or in real time.

Their results indicate that micro-targeting can lower advertising costs and raise incremental profit even in the face of ad avoidance. For example, with advance buy, it is possible to lower costs per target ad view by over 90%. With real-time buy, it is further possible to lower target costs per view while concurrently increasing target views; in one schedule, views to the target households were increased by 47% while costs concurrently reduced by 7%.

Overall, they find that the greatest potential to increase the profitability of advertising arises from (1) the integration of purchase data with viewing data (single-source data) and (2) the ability to buy placements real-time instead of in advance.

As single-source data becoming increasingly available and with the development of an advertising ecosystem to enable real-time buying on TV, their analysis highlights the
importance of these advances and a means of exploiting them. More-effective targeting may open new paths to TV advertising pricing by allowing TV networks and cable companies to sell “boxes” as well as shows to advertisers.

Yiting Deng is an Assistant Professor at the UCL School of Management, University College London. Carl F. Mela is the T. Austin Finch Foundation Professor of Business Administration at the Fuqua School of Business, Duke University and MSI 2017-2019 Executive Director. A portion of this work was completed while the second author was visiting at the Rotterdam School of Management, Erasmus University.

Acknowledgments
The authors would like to thank Bryan Bollinger, Pedro Gardete, Roni Shachar, Richard Staelin, Peng Sun, Kenneth Wilbur, Yi (Daniel) Xu, participants at the 2016 Marketing Science Conference, and seminar participants at Cornell University, Duke University, Erasmus University, HKUST, McGill University, New York University, Peking University, Syracuse University, University at Buffalo, University of California-Riverside, University College London, University of Miami, University of Notre Dame, and University of Rochester for useful comments. The usual disclaimer applies.
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 (2016)), American adults spend about 4 hours and 39 minutes watching traditional TV each day, far exceeding the combined daily consumption of PC, smartphone and tablets. Despite fierce competition from digital advertising, TV advertising is still huge and growing. eMarketer forecasts that TV spending will be $71.29 billion by the end of 2016 (eMarketer (2016)). One of the key advantages of Internet advertising has been its ability to target households. However, with the proliferation of digital set-top boxes (STBs), video recorders (DVRs) that enable advertisers to target households, this is changing. It is estimated that the number of US households reachable by addressable TV ads is now 68 million. Annual revenues for addressable TV advertising were $300 million in 2015 with an astonishing 100% year over year growth in revenues (Urbanski (2016)).

The digital targeting of advertising on television is transformative in two regards. First, digital TV affords household-level measures of TV viewing, making it possible to better forecast household advertising exposure.\textsuperscript{1} Second, set-top boxes enable household-show level targeting, enhancing targeting precision. While improving the cost efficiency of advertising exposures is of interest in its own right, coupling set-top viewing data with purchase data can also be used to evaluate the profitability of campaigns at the household-show level. 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 profitability.

Several companies have already started offering targeted TV advertising solutions. 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 return on investment. 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. Axiom

\textsuperscript{1}Concerns about measurement (e.g., ad viewability, the nature of content surrounding the ad, and the validity of exposures) are substantially attenuated in the TV market relative to display advertising online.
TV, launched in January of 2016, enables advertisers to combine first and third party data with TV viewing information to target advertising on TV.

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 (average = 85% of days) and for most of the evening (average = 88% of prime-time hours). Thus, any gains from targeting largely arise from what a viewer watches rather than whether they watch. Once a viewing session commences, 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 returns to household-level advertising targeting, and that targeting on genres and time will have smaller effects on advertising avoidance.

Using these insights to develop an integrated model of viewing predicated on a single consumption utility framework, we capture the show sampling, show viewing, and show exit processes described above. The approach builds upon Arcidiacono et al. (2015)’s continuous time dynamic discrete choice models, which we extend by incorporating sampling/consideration behavior. 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 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 advertising exposures or to maximize its incremental profit from advertising; and 2) whether the advertising purchase is made in advance or in real-time. The advantage of cost-based targeting is that it is predicated upon only advertising viewership, and does not require a model of sales response to advertising. However, cost-based targeting ignores the link between advertising and sales, and thus the

---

2We use the terms “household” and “viewer” interchangeably.
profitability of a campaign. 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 ad view by over 90%. With real time buy, it is further possible to lower target costs per view while concurrently increasing target views; in one schedule views to the target households can be increased by 47% while concurrently reducing costs by 7%. In short, we document dramatic increases in the cost efficiency of targeted media buys. Likewise, we find that advertisers can improve their advertising profitability when real-time ad buys are enabled. The effect is smaller for advance buys because uncertainty about advertising exposures attenuates advertising response, making advertising less profitable.

The remainder of the 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 describes the estimation results. Based on these results and a purchase model, §6 conducts counterfactual policy experiments and evaluates potential gain to be realized from targeting. Finally, we conclude with a schedule of next steps in §7.

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. (Forthcoming)). 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 (2015)). 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. (2016), who explore implications of individual-level targeting to consumers who are less likely to avoid 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 set-top box usage data, purchase data, advertising data, and programming data, as summarized in Figure 1. This combination is often called single source data because it covers the entire TV viewing and purchase experience for a set of households. The set-top box data (TiVo log files) track each household’s complete usage of a TiVo set-top box 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 set-top box data, advertising data and programming data will be used to estimate the viewing model (§4-§5). The purchase data and the Nielsen data will be used along with viewing model estimates in policy experiments on targeting (§6). 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 set-top box. They record every keystroke of the set-top box 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.

\[\text{For a more detailed description of the field study and the data, please see Bronnenberg et al. (2010).}\]
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 2006.4

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.

4The 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, Creative Execution, and Prices)

The TNS advertising schedule data describe advertising schedules for 61 national broadcast and cable TV networks. For each advertisement, the data report the precise air time, network, length, attributes of the advertised product (e.g., product category, company and brand), name and genre of the associated show, location of the commercial break within the show (i.e., pod) and the slot 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 network of the advertisement, and assessing whether or not the network was viewed at that time.

The data also contain a brief description of the advertisement, which is a summary of the creative execution for the advertisement (e.g., “mega roll/bear changes role” for Charmin bathroom tissue). A creative execution is often run for several weeks and is used to analyze advertising response. In §6, we will consider the effect of dropping a particular advertising creative on sales.

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 broadcast 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 observation in the TiVo Log Files is tagged with a unique TMS identifier, which is used to match with the TMS data, resulting in a description of each show viewed.
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 conditional 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 2 shows most households watch TV most evenings: on average, households watch prime-time TV on 85% of the days in the sample. Figure 3 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 2: Fraction of Days With 8 p.m.-Midnight Viewing (by Household, n = 834). 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.](image)

3.2.2 Show Sampling and Viewing

When starting to watch TV, a household first chooses which show to watch.\(^5\) Due to incomplete knowledge of program and episode quality, households sample shows (either live or recorded) for a brief duration.

\(^5\)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.
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: Frequency of Cumulative Daily TV Usage, 8 p.m.-Midnight (by Household-Day, $n = 304,410$). 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 4 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.\(^6\)

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 5).\(^7\) The hazard rate decreases drastically within the first 3 minutes, and remains

\(^{6}\)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.

\(^{7}\)The show switching spikes in Figure 5 are coincident with the starting times of shows on other networks. 69.2% of all shows start on the hour, 26.9% start on the half hour, comprising the bulk of starting times. However, shows also commence
Figure 4: Show Completion Rate (by Household-Show, \( n = 6,617,351 \)). 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%.

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).\(^8\) Based on this categorization, Figure 6 indicates that, in 70% of the cases, a household decides to watch the first show sampled. Hence, sampling is informative of preferences.

Figure 5: Hazard Rate for One-hour Shows Watched from the Broadcast Start Time (across Household-Show, \( n = 611,571 \))

\(^8\)A sensitivity analysis on sampling the threshold duration using alternative lower bounds of 10 seconds, 20 seconds and 60 seconds indicates the number of shows sampled vary less than \( \pm 7\% \) relative to the current 30 second threshold.
Figure 6: Number of Shows Sampled Before Watching One (by Household-Show, \(n = 2,899,749\)). The x-axis is the number of samplings before watching a show, and the y-axis is the percent of the particular number in all observations.

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 set-top box’s storage capacity was able to store 40 hours of programming. Set-top box 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 therefore informative about relative show preferences.

### 3.2.4 Advertising Viewing

Advertising viewership is predicated on the series of decisions shown in Figure 7. First, advertising viewing requires one to be watching the show when the advertisement airs (a viewer is *exposed*). Second, a household must decide whether to watch a show live or recorded. 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 avoidance occurs when a fast-forward starts before the advertisement and ends after it. Partial avoidance occurs when the household starts or stops fast-forwarding (i.e., zipping) during the advertisement and/or switches channels into or out of the advertisement (i.e., zapping). As expected, recorded shows are more subject to avoidance, as evidenced by the nearly 80% avoidance rate (mostly zipping) for recorded shows and a lower than 15% avoidance rate (mostly zapping) for live or near-live shows.

Figure 7: Advertising Viewing and Avoidance

To further exemplify avoidance patterns in recorded shows, we depict in Figure 8 the timing of zips and commercial breaks during a one-hour-long recorded episode of “CSI: Crime Scene Investigation” on December 8, 2005. Figure 8 indicates zips coincide with commercial breaks. Similar patterns exist in other shows. 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.

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

### 3.2.5 Advertising Costs

Table 2 and Figure 9 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.

To ascertain whether advertising prices relate to viewership, we regress the advertisements’ prices (in $1,000) on ratings and network dummies. Results indicate a positive and significant relationship between
<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>Prior ad avoided</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>

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>0.011</td>
<td>0.010</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>

price and rating \( (est = 7.8, se = 0.9) \), and the network coefficient is the highest for FOX \( (est = 32.9, se = 3.6) \), followed in turn by CBS \( (est = 25.3, se = 4.7) \), NBC \( (est = 16.8, se = 4.0) \) and ABC \( (est = 12.7, se = 3.6) \). Presumably the differences across networks relate to the demographics of the shows viewed (e.g., Goettler (2012)).

4 TV Viewing Model

4.1 Summary of Model Components

This summary overviews the various model components used in our targeting analysis. First, we consider the television and advertising viewing model in §4. This modeling component enables us to predict advertising views. Second, we consider the advertising response model in §6.1. This modeling component enables us to link advertising views to sales. Combined, the first two model components enable us to ascertain how changes in advertising buys affect sales. Third, using the link between advertising response and costs, we conduct counterfactual targeting analyses in order to improve the efficiency of media buys.
Figure 9: Distribution of Per-exposure Advertising Price Based on Nielsen Ratings (July 2006)

This modeling component is discussed in §6.2. Figure 10 overviews these model components, enumerates data requirements, and indexes the section where each modeling component is discussed. We begin our discussion with the viewing model.

Figure 10: Model Overview

4.2 The Viewing Model Overview

The TV viewing model comprises three components motivated by the preceding empirical discussion: TV show sampling and watching (§4.4), TV show recording (§4.5), and advertising viewing (§4.6). All
three components are predicated upon the theoretical concept of flow utility, that is, the moment-by-
moment consumption benefit a viewer derives from watching a TV show or advertisement, and non-TV
activities (the outside good). In this flow utility framework, viewers 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, viewers 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 after choosing a show to watch, the flow utility of the show changes (for example a
new scene or show segment), and the viewer 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 sessions 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 (show, advertisement, and outside good) in §4.3. Based on
these flow utilities, we then describe the viewer choice process (show sampling, show viewing, show
recording, and advertising viewing) in §4.4, §4.5 and §4.6 that follow.

4.3 Flow Utilities

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

4.3.1 Show Utility

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

\[
\begin{align*}
\hspace{1cm} u_{itn}^S &= X_{itn}^S \beta_i^S + v_{itn}^S + \varepsilon_{itn}^S \\
\equiv & \hspace{1cm} \bar{u}_{itn}^S + v_{itn}^S + \varepsilon_{itn}^S 
\end{align*}
\]
where $X_{itn}^S$ is a vector that captures show characteristics and viewer $i$’s past viewing behavior, including genre, network, show length, the percent of show aired when sampling begins (i.e., viewing offset), the number of previous episodes of the program that the viewer has sampled in the preceding week, and whether the viewer was watching network $n$ before the current choice occasion. The last two capture state dependence in viewing behavior. $\nu_{itn}^S$ represents a viewer-show specific error term observed by the viewer but not by the researcher. $\epsilon_{itn}^S$ represents viewer uncertainty pertaining to program and episode quality prior to sampling that is revealed to the viewer only after sampling the show. We assume $\nu_{itn}^S$ is i.i.d. standard Type I Extreme Value distributed, and $\epsilon_{itn}^S$ is i.i.d. Type I Extreme Value distributed with mean zero.\textsuperscript{11}

4.3.2 Advertisement Utility

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

$$u_{itn}^A = X_{itn}^A \beta_i^A + \epsilon_{itn}^A$$

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

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

4.3.3 Outside Good Utility

When allocating time, the viewer contrasts the utility from viewing TV to the best available alternative (i.e., the outside good). If the utility from viewing TV exceeds that of the outside good, the viewer will

\textsuperscript{11}To accommodate the potential that viewer experience might reduce this uncertainty prior to sampling, we interact past viewing with the remaining model covariates. Online Appendix A shows that this approach is equivalent to allowing for heterogeneous uncertainty based on differential familiarity with shows. We thank an anonymous reviewer for motivating this analysis.
watch TV. As such, the flow utility of the outside good is tantamount to the “opportunity cost” of time, posited to vary by viewer \((i)\) and time \((t)\), and denoted as:

\[
\begin{align*}
    u^O_{it} &= X^O_{it} \beta_i^O + \nu^O_{it} \\
    &\equiv \bar{u}^O_{it} + \nu^O_{it},
\end{align*}
\]

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

### 4.4 TV Show Viewing

Owing to \(\varepsilon^S_{itn}\) in Equation (1), viewers face \textit{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 viewer first samples the alternative with the highest \textit{ex-ante} expected viewing utility given the observed show characteristics and \(\nu^S_{itn}\), the shock unobserved to the researcher. This ordering of alternatives is simple for the viewers to do (suggesting the process is a parsimonious representation) and is also consistent with the optimal search ordering implied by Weitzman (1979) and Kim et al. (2010).\(^{12}\) After the short exposure to the show, the viewer observes the utility shock for that show, \(\varepsilon^S_{itn}\), 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 \(\varepsilon^S_{itn'}\) at the other shows \(tn'\) is still unknown to the viewer). If the flow utility of the current show is lower, the viewer samples other shows.

If the flow utility of the current show is higher than the expected best remaining alternative, the viewer selects the show and enters a “flow” state of watching. During this flow state, the flow utility of the show

\(^{12}\)Weitzman (1979) shows that, in sequential search, the search order follows the order of reservation utility. Consumers first search for products with the highest reservation utility \(z_{ij}\) among the products not yet searched. Kim et al. (2010) show that the reservation utility is given by \(z_{ij} = V_{ij} + \zeta \left( \frac{c_{ij}}{\sigma_{ij}} \right) \times \sigma_{ij}\) when the utility specification is \(u_{ij} = V_{ij} + e_{ij}, e_{ij} \sim N \left( 0, \sigma^2_{ij} \right)\). \(c_{ij}\) is the search cost, and \(\zeta \left( \frac{c_{ij}}{\sigma_{ij}} \right)\) is a scalar function that translates standardized search cost \(\frac{c_{ij}}{\sigma_{ij}}\) into a multiplier on \(\sigma_{ij}\). In the TV sampling context, search cost is negligible because sampling is trivially done and \(\sigma_{ij}\) is equal across shows. Hence the equivalent “reservation utility” is monotone in \(\bar{u}^S_{itn} + \nu^S_{itn}\). Therefore, under these assumptions, viewers start search with the option having the highest \(\bar{u}^S_{itn} + \nu^S_{itn}\).
remains constant until an external “shock” randomly arrives that changes the flow utility by perturbing $\varepsilon_{i|n}^S$. 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 viewer 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 viewer switches. An analogous process holds for advertising. When an advertisement appears in the midst of the show, the flow utility changes to that of the advertisement, and the viewer again compares the flow utility to other options. Figure 11 overviews this process and the sections of the paper that elaborate on the show sampling and viewing decisions (while the ad viewing process is described in §4.6).

![Decision Making in One Viewing Session](image)

Figure 11: Decision Making in One Viewing Session

We elaborate on the sampling and viewing processes next.

### 4.4.1 Sampling

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

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

(4)

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

After sampling show $tn$, viewer $i$ observes $e_{itn}^S$ and therefore $u_{itn}^S$. The viewer 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:

$$IV_{it, \mathcal{N}_it} \mid \{v_{itn}^S\}_n, v_{it}^O = E\{e_{itn}^S\}_{n' \in \mathcal{N}_it} \max_{n' \in \mathcal{N}_it} \{\bar{u}_{itn'}^S + v_{itn'}^S + e_{itn'}^S, \bar{u}_{it}^O + v_{it}^O\},$$

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

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

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

$$Pr\left(y_{itn}^W = 1 \mid y_{itn}^S = 1, \{v_{itn}^S\}_n, v_{it}^O\right) = Pr\left(u_{itn}^S \geq IV_{it, \mathcal{N}_it} \mid \{v_{itn}^S\}_n, v_{it}^O\right) = 1 - F_{\epsilon_{itn}^S}(\cdot),$$

(5)

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

Computation of $Pr\left(y_{itn}^W = 1 \mid y_{itn}^S = 1\right)$ involves integrating out $\{v_{itn}^S\}_n$ and $v_{it}^O$ in $Pr\left(y_{itn}^W = 1 \mid y_{itn}^S = 1, \{v_{itn}^S\}_n, v_{it}^O\right)$. The sampling order implies $v_{itn}^S$ and $v_{itn}^O$ in Equation (5) are truncated respectively below $\bar{u}_{itn}^S + v_{itn}^S - \bar{u}_{itn'}^S$ and below $\bar{u}_{itn}^S + v_{itn}^S - \bar{u}_{it'}^O$. 

---

13For 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 \( u^{S}_{itn} < IV_{it \setminus n} \), viewer \( i \) samples another show if the inclusive value of remaining shows is higher than the value of the outside good. The choice set is now \( \tilde{N}_{it} = N_{it} \setminus n \), and the probability of sampling show \( t \tilde{n} (\tilde{n} \in \tilde{N}_{it}) \) is:

\[
Pr(y_{it \tilde{n}} = 1) = \frac{\exp(\tilde{u}^{S}_{it \tilde{n}})}{\exp(\tilde{u}^{O}_{it}) + \sum_{n' \in \tilde{N}_{it}} \exp(\tilde{u}^{S}_{itn'})}.
\]

(6)

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

4.4.2 Watching

Upon selecting show \( tn \) to view, viewer \( i \) obtains viewing flow utility \( u^{S}_{itn} \) until the show ends or the viewer exits viewing, whichever comes first. The decision to stop watching is driven by arrival of external shocks that change the flow utility (Arcidiacono et al. (2015), Nevskaya and Albuquerque (2013)). If the external shock is sufficiently negative, the viewer terminates the show.\(^{14}\)

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

To characterize the duration until the arrival of a new external shock, we assume a homogeneous Poisson process with rate \( \lambda_{itn} \) for viewer \( i \) and show \( tn \).\(^{15}\) \( \lambda_{itn} \) is parameterized as a function of genre:

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

(7)

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.

\(^{14}\)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.

\(^{15}\)To be clear, there are two error processes. One is a logit error related to the unobserved (to the researcher) show flow utility at time \( t (\varepsilon^{S}_{itn}) \). The other is a Poisson process that defines a potential changepoint (at \( t' > t \) wherein the unobserved show flow utility error can change (e.g., from \( \varepsilon^{S}_{itn} \) to \( \varepsilon^{S}_{itn'} \)). This changepoint is unobserved to the researcher but known to the viewer (such as a change in stories while watching the news). Together, these processes imply that the new flow utility shock is still drawn from the Extreme Value distribution, but that the timing of these change points follows a Poisson process.
Under this Poisson assumption, the probability of viewer $i$ exiting show $tn$ at time $t'$, $q_{itnt'}$, is given by the probability that the flow utility upon receiving a new viewing shock falls below the alternative options:

$$q_{itnt'} \mid \{v_{itn}^S\}_n, v_{it}^O = Pr \left( \bar{u}_{itn}^S + v_{itn}^S + \epsilon_{itn}' < \max \left\{ \bar{u}_{it}^O + v_{it}^O, \ln \left( \sum_{n' \in N_{it}'} \exp \left( \bar{u}_{itn'} + v_{itn'} \right) \right) \right\} \right)$$

$$= F_{\epsilon_{itn}'} \left( \max \left\{ \bar{u}_{it}^O + v_{it}^O, \ln \left( \sum_{n' \in N_{it}'} \exp \left( \bar{u}_{itn'} + v_{itn'} \right) \right) \right\} - \bar{u}_{itn} - v_{itn} \right).$$

Computation of $q_{itnt'}$ involves integrating out $\{v_{itn}^S\}_n$ and $v_{it}^O$ in $q_{itnt'}$, and $q_{itnt'}$ 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 $n'$, $\bar{u}_{itn'} + v_{itn'}$ changes and $q_{itnt'}$ would change accordingly. Hence, $q_{itnt'}$ is piecewise constant and changes whenever shows on alternative networks ($N_{it}$) change. Note that in between shocks, $q_{itnt'}$ remains fixed.

Using an approach similar to Arcidiacono et al. (2015) and Nevskaya and Albuquerque (2013), Appendix B shows for $q_{itnt'}$ that is piecewise constant with segment $1, \ldots, M$, the CDF of the viewing length $l_{itn}^*$ is:

$$F_{l_{itn}^*} (\bar{t}) = Pr \{l_{itn}^* \leq \bar{t}\} = 1 - e^{-\lambda_{itn} \sum_{m=1}^{M} l_{itn}^m q_{itn}^m},$$

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

The probability density function of $l_{itn}^*$ is therefore:

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

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

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

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

Marketing Science Institute Working Paper Series
where $l^{W}_{itn}$ is the time viewer $i$ spends watching show $tn$, and $L_{itn}$ is the remaining length of show $tn$ when sampled.

### 4.5 TV Show Recording

The TiVo set-top box 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 set-top box, 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 viewer $i$ records show $tn$ is given by:

$$Pr(\gamma_{itn}^R = 1) = \frac{\exp(\bar{u}_{itn}^S)}{\sum_{n'} \exp(\bar{u}_{itn'}^S)},$$

where $\gamma_{itn}^R$ is an indicator that show $tn$ is recorded by viewer $i$.

### 4.6 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 viewer’s alternative set includes the opportunity cost of time. Below, we formalize these points.

#### 4.6.1 Zapping

In the viewing model introduced above, the viewer can choose to avoid advertisements in a live show by channel switching (zapping). Similar to §4.4.2 where viewers can make channel switching decisions during program content upon receiving an external shock, viewers make zapping decisions during each advertisement. Hence, this component of the viewing model is analogous to the 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:

\[
Pr\left( y_{itn}^{AL} = 0 \mid \{v_{itn}^{S}\}_{n}, v_{it}^{O} \right) = Pr\left( \bar{u}_{itn}^{A} + \epsilon_{itn}^{A} < \max \left\{ \bar{u}_{it}^{O} + v_{it}^{O}, \ln \left( \sum_{n' \in \mathcal{N}_{it}} \exp \left( \bar{u}_{itn'}^{S} + v_{itn'}^{S} \right) \right) \right\} - c_{i} \right) = F_{\epsilon_{itn}^{A}} \left( \max \left\{ \bar{u}_{it}^{O} + v_{it}^{O}, \ln \left( \sum_{n' \in \mathcal{N}_{it}} \exp \left( \bar{u}_{itn'}^{S} + v_{itn'}^{S} \right) \right) \right\} - c_{i} - \bar{u}_{itn}^{A} \right),
\]

where \(y_{itn}^{AL}\) is an indicator that the live advertisement \(tn\) is viewed (not zapped) by viewer \(i\), \(c_{i}\) is the zapping cost faced by viewer \(i\).\(^{16}\) Computation of \(Pr\left( y_{itn}^{AL} = 0 \right)\) involves integrating out \(\{v_{itn}^{S}\}_{n}\) and \(v_{it}^{O}\) in \(Pr\left( y_{itn}^{AL} = 0 \mid \{v_{itn}^{S}\}_{n}, v_{it}^{O} \right)\).

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

\[
Pr\left( y_{itn}^{AR} = 0 \right) = \frac{\exp \left( \bar{u}_{it}^{O} \right)}{\exp \left( \bar{u}_{it}^{O} \right) + \exp \left( \bar{u}_{itn}^{A} \right)},
\]

where \(y_{itn}^{AR}\) is an indicator that the recorded advertisement \(tn\) is viewed (not zipped) by viewer \(i\).\(^{17}^{18}\) 

---

\(^{16}\)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 viewer 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.

\(^{17}\)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.

\(^{18}\)We also investigated whether advertising viewing decision depends on past purchase. Because we only have purchase information for CPG categories, we re-estimated this advertising viewing model by adding the purchase history of the advertised brand as an additional covariate in advertising viewing flow utility. We considered 4 different measures of purchase history: 1) whether the viewer has purchased the brand before; 2) whether the viewer has purchased the brand in the previous 7 days; 3) whether the viewer has purchased the category at least once before in the panel; and 4) whether the viewer has purchased the category in the previous 7 days. None of these measures resulted in a significant effect of purchase on ad viewing. The corresponding estimates and p-values for the four purchase variables are 0.074 (p=0.40), –0.028 (p=0.85), 0.043 (p=0.76) and –0.007 (p=0.95). Therefore, past history in neither the product category nor the brand seems to affect advertising viewing decision in our data. As such we report the more parsimonious specification in the paper. We thank an anonymous reviewer for prompting this analysis.
The observed show sampling and watching decisions, show recording decisions and advertising viewing decisions enable us to recover flow utilities of TV shows, TV advertisements, and the outside good. We discuss estimation in the next section.

5 Viewing Model Estimation and Results

5.1 Estimation

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

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

As virtually every viewer watches only a handful of available networks, we construct viewer-specific consideration sets based on viewing history. For each viewer, the consideration set of networks consists of the smallest number of networks that collectively account for at least 90% of prime-time viewing time. On each viewing occasion, the choice set comprises the following two types of shows: i) live shows that are available on networks within the consideration set; and ii) shows stored on set-top box 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 advertising data. In an initial effort to capture the effects

---

19 On average, there are 1,313 sampling occasions and 1,993 advertising viewing occasions per viewer.
20 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 D.
of product category on advertising preference, we classify the 574 product categories into four general categories: consumer packaged goods (CPG), services, pharmaceuticals, and other goods.

5.2 Results

Table 3 reports the estimation results of the viewing model. The second column reports the median (across viewers) of parameter estimates in flow utilities of shows as well as the percentages of viewers with significant positive and negative estimates (5% level).\(^{21}\) For most viewers, shorter shows are associated with higher utility as indicated by 65.7% of people having a significant negative coefficient on “length”. As expected, the average viewer prefers familiar shows, indicating positive state dependence (“loyalty”) in show utility. This result is consistent with Schweidel and Moe (2016), who find that increased viewing of past episodes from the same series leads to more series viewing in the future. Also consistent with prior literature (e.g., Moshkin and Shachar (2002); Esteves-Sorenson and Perretti (2012)), state dependence is evidenced in network viewing as indicated by the generally positive coefficients on “lag network”. The coefficients for “live” indicate the average viewer prefers live shows over recorded shows. The negative coefficient for “viewing offset” (which measures how far into a show a viewer starts watching) implies shows with less elapsed time from the start of the show is preferred. In other words, viewers prefer to watch shows on which they have missed less content.

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

\(^{21}\)As noted in Footnote 11, our estimation model includes interaction terms between loyalty and other covariates to control for show familiarity. Because these are control variables and most are insignificant, we do not report them but detailed estimates are available from the authors. We further note that the insights in the paper are largely invariant to the inclusion of these interactions.
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>−0.20 (1.2%, 65.7%)</td>
<td>5.23 (66.6%, 0.6%)</td>
</tr>
<tr>
<td>Episodes sampled in preceding week</td>
<td>0.83 (49.5%, 15.9%)</td>
<td>−0.11 (6.4%, 16.3%)</td>
</tr>
<tr>
<td>Lag network</td>
<td>0.63 (67.6%, 1.9%)</td>
<td>0.05 (11.3%, 10.8%)</td>
</tr>
<tr>
<td>Live</td>
<td>9.79 (79.9%, 0.0%)</td>
<td>1.51 (55.3%, 1.0%)</td>
</tr>
<tr>
<td>Viewing offset</td>
<td>−4.96 (0.0%, 82.0%)</td>
<td>−0.11 (3.8%, 7.3%)</td>
</tr>
<tr>
<td>Prior ad viewed</td>
<td>−0.21 (12.9%, 34.2%)</td>
<td>−0.77 (9.0%, 25.9%)</td>
</tr>
<tr>
<td>First ad break</td>
<td>−0.47 (8.2%, 48.2%)</td>
<td>−0.10 (8.8%, 15.2%)</td>
</tr>
<tr>
<td>Last ad break</td>
<td>−0.74 (6.4%, 48.4%)</td>
<td>−0.78 (8.6%, 24.7%)</td>
</tr>
<tr>
<td>First slot in a break</td>
<td>0.26 (34.7%, 15.9%)</td>
<td>0.30 (7.6%, 10.2%)</td>
</tr>
<tr>
<td>Last slot in a break</td>
<td>0.23 (32.3%, 12.8%)</td>
<td>−0.10 (7.8%, 11.4%)</td>
</tr>
<tr>
<td>Genre: Drama</td>
<td>−0.21 (12.9%, 34.2%)</td>
<td>−0.77 (9.0%, 25.9%)</td>
</tr>
<tr>
<td>Genre: Comedy</td>
<td>−0.47 (8.2%, 48.2%)</td>
<td>−0.10 (8.8%, 15.2%)</td>
</tr>
<tr>
<td>Genre: Reality TV</td>
<td>−0.74 (6.4%, 48.4%)</td>
<td>−0.78 (8.6%, 24.7%)</td>
</tr>
<tr>
<td>Genre: Talk shows</td>
<td>0.26 (34.7%, 15.9%)</td>
<td>0.30 (7.6%, 10.2%)</td>
</tr>
<tr>
<td>Genre: News</td>
<td>0.23 (32.3%, 12.8%)</td>
<td>−0.10 (7.8%, 11.4%)</td>
</tr>
<tr>
<td>Genre: Sports</td>
<td>−0.25 (13.7%, 24.6%)</td>
<td>−0.15 (7.3%, 15.2%)</td>
</tr>
<tr>
<td>Network: ABC</td>
<td>0.81 (45.6%, 11.5%)</td>
<td>−0.03 (12.5%, 16.7%)</td>
</tr>
<tr>
<td>Network: CBS</td>
<td>0.33 (33.6%, 20.0%)</td>
<td>−0.14 (10.9%, 15.3%)</td>
</tr>
<tr>
<td>Network: NBC</td>
<td>0.50 (40.3%, 19.3%)</td>
<td>−0.18 (11.2%, 17.5%)</td>
</tr>
<tr>
<td>Network: FOX</td>
<td>0.24 (24.9%, 15.2%)</td>
<td>−0.14 (10.0%, 12.9%)</td>
</tr>
<tr>
<td>Network: USA</td>
<td>−0.09 (9.0%, 10.8%)</td>
<td>−0.22 (6.0%, 6.8%)</td>
</tr>
<tr>
<td>Network: Comedy Central</td>
<td>−0.12 (8.4%, 8.2%)</td>
<td>0.78 (4.4%, 2.0%)</td>
</tr>
<tr>
<td>Product category: CPG</td>
<td>−0.02 (4.0%, 4.2%)</td>
<td>0.01 (4.8%, 3.6%)</td>
</tr>
<tr>
<td>Product category: service</td>
<td>−0.02 (4.0%, 4.2%)</td>
<td>0.01 (4.8%, 3.6%)</td>
</tr>
<tr>
<td>Product category: drug</td>
<td>−0.09 (2.0%, 5.3%)</td>
<td>−0.22 (6.0%, 6.8%)</td>
</tr>
</tbody>
</table>

There also exists extensive heterogeneity in zapping cost across viewers, meaning some viewers do not avoid live advertisements (and presumably are better targets than those that do). The mean, median and the standard deviation of the zapping cost are respectively 2.7, 0.9 and 4.7, and the 2.5% and 97.5% quantiles are respectively 0.01 and 2.9.

Finally, 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 viewers in viewing preferences for genre, network, and advertising. Heterogeneity in viewing preferences, together with heterogeneity in advertising response, suggests the potential gains available from advertising targeting.

6 Policy Experiments: Advertising Targeting

This section conducts advertising targeting policy experiments. We first link advertising viewing with sales response in §6.1. Predicated on viewing behavior and advertising response, we discuss various targeting strategies in §6.2.

6.1 Advertising Response

To measure advertising response, we select seven product categories evidencing high variation in advertising and sales: children’s yogurt, children’s cereal, regular cola, diet cola, sports drink, toothpaste, and bathroom tissue. These categories are regularly purchased and frequently advertised, with high cross-sectional and temporal variation in both purchase and advertising. Within these categories, we consider 22 leading brands that have non-negligible unit market share and that advertised during the sample period.\(^\text{22}\)

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

\[
U_{ijm}^p = Z_{ijm} \theta_{ic} + h_{ic} (A_{ijm}) + \varepsilon_{ijm}^p, \tag{15}
\]

where \(Z_{ijm}\) is a vector that includes a brand fixed effect, the price of brand \(j\) at shopping trip \(m\), an indicator of whether brand \(j\) is on promotion (either display or feature) at shopping trip \(m\),\(^\text{23}\) and an indicator

\(^{22}\)For more detail on these purchase data refer to Online Appendix E.

\(^{23}\)Price and promotion are computed as weighted averages of item-level price and promotion, with the weight being store-level unit sales.
of whether household $i$ purchased brand $j$ in the previous category purchase. The function $h_{ic}(A_{ijm})$ captures the advertising effect.\textsuperscript{24} Finally, $\varepsilon_{ijm}^p$ is an idiosyncratic error term affecting the inherent valuation of brand $j$ at shopping trip $m$, and it is observed by the household but not by the researcher.

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

$$h_{ic}(A_{ijm}) = \sum_{a=1}^{A} N^A_{ijam} \gamma_{jac} + \left( N^A_{ijm} - \sum_{a=1}^{A} N^A_{ijam} \right) \gamma_{j0c}, \quad (16)$$

where $a \in \{1, \ldots, A\}$ indexes advertising creatives (as defined in §3.1.3), and $N^A_{ijam}$ is household $i$’s number of views on advertising creative $a$ since the previous category purchase or in the past 7 days, whichever is smaller.\textsuperscript{25} We combine creatives with smaller number of views as they are individually likely to have a negligible effect on demand. In particular, if $N^A_{ijm}$ is household $i$’s total number of advertising views across brand $j$’s creatives during this window, then $N^A_{ijm} - \sum_{a=1}^{A} N^A_{ijam}$ represents the total number of views on all other creatives with smaller number of views. Equation (16) allows us to measure creative-specific effects for major creatives (e.g., Homer and Yoon (1992); Malaviya et al. (1996); Vakratsas and Ambler (1999); Tellis et al. (2005)) while pooling the effect of the smaller creatives.

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

$$U_{i0m}^p = \varepsilon_{i0m}^p. \quad (17)$$

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

$$Pr(\gamma_{ijm}^p = 1) = \frac{\exp \left( Z_{ijm} \theta_{ic} + h_{ic}(A_{ijm}) \right)}{1 + \sum_{j' = 1}^{J} \exp \left( Z_{ij'm} \theta_{ic} + h_{ic}(A_{ij'm}) \right)}. \quad (18)$$

\textsuperscript{24}Firms may target households that are more responsive to advertising. To address this potential endogeneity concern, we compute for each household and each brand: i) the brand’s share in the household’s category purchase, and ii) the brand’s share in the household’s category advertisement exposure. We do not find a correlation between the two variables, implying the advertising variables are unlikely to be endogenous. This lack of correlation might be a consequence of the current targeting practice, and implies the potential for improvement. In addition, the between-household variance of advertising exposure is lower than the within-household variance. For the 22 brands in the sample, the mean and median portions of the total variance between households are respectively 20.1% and 15.9%, with the lowest being 4.0% and the highest being 47.9%.

\textsuperscript{25}We use a 7 day window as this typically represents two shopping trips, enhancing the likelihood one can attribute a given view to a purchase decision on a given shopping trip, akin to last touch attribution.
Using a latent class model (Kamakura and Russell (1989)) to capture household heterogeneity, the probability that household \( i \) in segment \( k (k = 1, \ldots, K) \) chooses brand \( j \) in shopping trip \( m \) is:

\[
Pr(y_{ijm} = 1 | i \in k) = \frac{\exp(Z_{ijm}\theta_{kc} + h_{kc}(A_{ijm}))}{1 + \sum_{j=1}^{J} \exp(Z_{ij'm}\theta_{kc} + h_{kc}(A_{ij'm}))},
\]

(19)

where \( h_{kc}(A_{ijm}) \equiv \sum_{a=1}^{A} N_{ijam}^{A} k_{jac} + \left( N_{ijm}^{A} - \sum_{a=1}^{A} N_{ijam}^{A} \right) k_{j0c} \).

The model is separately estimated for each product category, resulting in a unique set of advertising coefficients across product categories, consumer segments and advertising creatives. To assess whether or not advertising affects sales, Figure 12 plots the histogram of the asymptotic t-statistics of the estimated advertising coefficients.\(^{26}\) As in prior research, advertising effects are statistically small, but when they are distinguishable from zero, they are mostly positive. Were advertising effect zero, the distribution of the asymptotic t-statistics would follow a standard normal distribution based (per the Central Limit Theorem). A one-sample Kolmogorov-Smirnov test rejects the hypothesis that the asymptotic t-statistics follow a standard normal distribution (p<0.07), indicating that advertising effects overall are statistically greater than zero.\(^{27}\) To assess the external validity of our findings, we compare the percent of estimated advertising effects that are significant at different p-values, and compare that percentage with the percentages reported by existing review papers on advertising effects. The results are reported in Table 5, and are generally in line with previous findings.

<table>
<thead>
<tr>
<th>Table 5: Percent of Significant Advertising Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>P=0.05</td>
</tr>
<tr>
<td>----------------------------------------</td>
</tr>
<tr>
<td>Eastlack and Rao (1989)</td>
</tr>
<tr>
<td>Lodish et al. (1995)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Hu et al. (2007)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>This study</td>
</tr>
</tbody>
</table>

To assess the magnitude of advertising effects that are significant (at the level of 10%), we simulate the effect of removing the associated advertising creative on sales. Inspired by Sethuraman et al. (2011) who find on average, the long-term advertising elasticity is twice as high as the short-term elasticity, we...

\(^{26}\)Detailed estimation results are available from the authors.

\(^{27}\)At the level of 5%, 19 out of 129 consumer segment - creative pairs are significantly positive at p=0.05 (15%), and 4 are significantly negative (3%). Thus, negative effects are slightly less than one might expect by chance and while positive effects are substantially larger.

Marketing Science Institute Working Paper Series
measure both short-term and long-term effect. The short-term effect is obtained by simulating contemporaneous changes in market share holding all else constant. The long-term effect takes into account not only the immediate effect of eliminating the advertising creative, but also the carry-over effect of altered choice through the purchase event feedback measure (whether the consumer purchased the brand in the previous purchase).  

Figure 13 depicts the results in terms of percentage changes in market share. The length of the black segment of each bar represents the short-term effect, and the total length of the bar represents the long-term effect. As shown in the figure, the change is negative for the majority (18 out of 22) of consumer segment - creative pairs. Among such pairs, the combined short-term and long-term changes in own market share vary from -5.5% to -0.4%, suggesting the long-term effects of an advertising execution are not insubstantial.

---

28 Specifically, we first compute the purchase probability for each brand in the consumer’s first observed shopping trip with and without an advertising creative, and simulate a purchase decision based on these probabilities. Then, using the purchase event feedback term (lag brand purchase), we compute consumers’ purchase probabilities in the second shopping trip. We continue this process until the last observed shopping trip thereby obtaining a new purchase sequence for each consumer. The entire process is repeated 100 times for each consumer, and the new market share is obtained by averaging the simulated market shares obtained under the 100 simulations. The effect of advertising is then the difference between simulated long-term effects with and without an advertising creative.
Overall, we conclude that the significant advertising effects are generally positive as in the previous literature (e.g., Sethuraman et al. (2011)). However, the effects vary across consumer segments and creatives, suggesting the potential to enhance the efficiency of advertising spend.²⁹

6.2 Targeting Approaches

Currently, national TV networks typically sell advertising inventory in advance by show. In the 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 commercial space by show and air date, the exact location of the advertisement’s placement within the show is determined at a later stage. Advertising prices vary with the number of 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

²⁹In Appendix E.2, we apply a “model-free” approach to measure advertising effects, which does not involve assumptions regarding functional form, advertising decay, etc. We also find small but positive advertising effects.
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 buys in upfront markets. Second, advertisements can be inserted real time, analogous to current practices in Internet advertising. In this case, the advertiser observes all information at the time immediately preceding the available advertising slot. This information set includes the show watched just before the ad airs (and at the time of the commercial break). Thus, in the real time case, the advertiser knows the TV is on and the consumer is watching a show when the slot appears, whereas in the advance buy case, these two decisions are only known up to the model’s ability to forecast them.

Several metrics can be employed when targeting. At the simplest level, one can either maximize target views for a given budget or minimize a budget for a given view. The advantage of these metrics is that they do not require a model of advertising response or sales data to implement. In our analysis, we focus on minimizing a budget for a given view. At a more complex level, one can consider the role of profits or revenues. At the cost of invoking an advertising response model and 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 section in which we detail the alternative policy experiments and results. Different cells utilize different model outputs. The upper left cell uses advertising viewing model; the upper right cell uses advertising viewing model and sales response model; the bottom cells use not only models in the respective cell above, but also the show viewing model. Hence they have the most substantial data and computational requirements.  

<table>
<thead>
<tr>
<th></th>
<th>Cost</th>
<th>Profit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real-time Buy</td>
<td>§6.2.1</td>
<td>§6.2.3</td>
</tr>
<tr>
<td>Advance Buy</td>
<td>§6.2.2</td>
<td>§6.2.4</td>
</tr>
</tbody>
</table>

We explore these policies using the leading brand of bathroom tissue, Charmin over the hold-out period of July 2006. Our analyses abstract away from competitive response (from the retailers, the

30In Online Appendix F, we also consider an aggregate-level targeting approach wherein advertisers can only buy shows and not users. This corresponds to current practice. Results indicate that leveraging household viewing information substantially increases advertising efficiency even in this aggregated case.

31We focus upon bathroom tissue as it has broad penetration. Within this category, we focus upon Charmin as it has the largest number of advertising views (share=46.8%) and the second largest sales (share=24.1%). As a result, there is sufficient
competitors or the networks), and assume a fixed (per-exposure) advertising price to a show. Thus, the findings are best interpreted as a marginal improvement in advertising holding all else fixed. While this assumption is reasonable when considering changes in purchases by a single advertiser, a more systematic change in policy would require an assessment of how advertising rates might change in response to new targeting capabilities, and how the advertiser’s competitors might adjust their own advertising schedules in response. While a fruitful area for additional research, we believe this analysis extends beyond the scope of this paper, and interpret our results in light of this caveat.

6.2.1 Cost-Based Real-time Buy

We first consider the potential for Charmin to lower its per-view cost (views differ from exposures inasmuch exposures need not be viewed). To do this, we first compute each targeted household’s observed average per-exposure advertising cost, by averaging the total observed advertising costs across all exposures to that household in the holdout data. Next, we compute the predicted cost per view by dividing the total observed advertising costs for a household by the predicted viewership. For example, if there are 2000 exposures and 1000 views to household $i$, and the placement costs $500$, then the average cost per-exposure (denoted $c_i$) is $500/2000 = 0.25$, and average cost per view (denoted $c_i^*$) is $500/1000 = 0.50$.

Denote the predicted cost for a particular show’s views on network $n$ at time $t$ to household $i$ as $c_{itn}^*$, which is the ratio of the view’s exposure cost and its predicted viewing probability. Intuitively, if $c_{itn}^* < c_i^*$, then a more efficient allocation of expenditures is feasible in terms of cost per view by reallocating ad dollars to show $tn$. Therefore, a parsimonious rule for the advertiser would be to set a level $K$, conditioned on the information available at time $t$, such that the advertiser purchases advertising slot in show $tn$ if $c_{itn}^* < Kc_i^*$. For example a rule of $K = 1$ implies the advertiser should buy all slots with a per-view cost that is lower than the average cost per view under the current schedule. Lower values of $K$ imply a rule that leads to higher levels of purchase efficiency, but at the cost of reach (there will be only a limited number of slots that meet an increasingly stringent cutoff). We implement this purchase rule viewer-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. In the profit-based simulations, we consider Charmin’s advertising creative “Mega roll/bear changes roll”, which has significant effect on one consumer segment (est=0.46, se=0.16) and insignificant effect on the other segment (est=0.05, se=0.20). As a result of the heterogeneity in advertising response, targeting should be more efficacious.

32In all targeting simulations, we take 50 parameter draws to predict viewing probabilities.
by-viewer; beginning at the first advertising slot in the period and ending when either a) the period has ended, or: b) both advertising views and expenditures under the new rule exceed those observed in the data.33

Figure 14 portrays the total advertising views and costs (across households) under different values of $K$. In the figure, we note that total advertising costs decrease as $K$ decreases. This is primarily because a lower value for $K$ means that Charmin is buying advertisements that have lower costs per view, and because fewer advertising slots are available that meet this criterion. The effect of decreasing $K$ on views is more complex. On the one hand, a decrease in $K$ means fewer slots are available that meet that criteria of lower cost per view (as $K$ goes to 0, there will be no advertising). This implies reduced views owing to reduced reach. On the other hand, there is the potential that lower cost per view can lead to more views because one can buy more views for a fixed budget (with sufficient inventory available). How these two opposing forces tradeoff is an empirical question.

Findings suggest that when $K$ is set below 0.6 (implying high cost efficiency constraint on ad buys), advertising views surpass those under the current schedule (690) even though overall advertising costs are below the current level ($8.77, or $12.71 CPM). For example, when $K = 0.4$, Charmin’s views are increased by 2% while its costs are decreased by 41%; when $K = 0.6$, Charmin’s views are increased by 47% while its costs are decreased by 7%. 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 views. To maintain the same number of views (690), the cost is even lower. For instance, when $K = 0.4$, the CPM is $7.38, 41\%$ lower than the current level; when $K = 0.6$, the CPM is $7.99, 37\%$ lower than the current level.

Of additional interest is the decomposition of real-time buy cost efficiency gains into i) gains from placing advertisements in shows that have lower cost per exposure (i.e., reducing cost) and ii) gains arising from targeting advertisements to those who are less likely to avoid them (i.e., increasing views). To answer this question we replace the model’s forecasted advertising avoidance (denoted Model A) with the average observed advertising avoidance rate across viewers (denoted Model B); we then use Model B to re-impute advertising costs and views under the simple buying rule. Setting $K = 0.4$, our first finding is that it is no longer the case that views increase and costs decrease under Model B (i.e., no heterogeneity in ad avoidance); instead, both increase. Second, the average costs per view in the observed data is 1.27

33A generalization of this purchase rule is to soften either the minimum advertising views or the maximum budget constraints. This exercise yields similar insights as described below.
cents, 0.73 cents under model A, and 0.94 cents under Model B. Thus, show placement alone yields a 26% improvement in cost efficiency over the observed schedule, but coupled with the advertisement viewing model there is a 43% improvement. In other words, roughly 2/5 of the improvement in efficiency is due to reducing advertising avoidance and the other 3/5 is due to cheaper impressions.

![Figure 14: Advertising Views and Costs Under Different Purchase Thresholds ($K$). The current observed cost is $8.77 with 690 views and 780 exposures.](image)

### 6.2.2 Cost-Based Advance Buy

In this targeting scenario, the advertiser purchases advertising slots in advance for a given period, minimizing the total cost of expected views for targeted households. In other words, advertisers are unable to condition on current viewing, meaning that they need to predict show viewing as well as advertising viewing.\(^{34}\)\(^{35}\)

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

$$
\text{Min} \{x_{i tn}\}_{i,n} \quad \text{subject to} \quad \sum_t \sum_n c_{tn} x_{i tn}^* \\
\quad s.t. \quad x_{i tn}^* \in \{0, 1\}, \forall t, n,
$$

\(^{34}\)Because almost all recorded shows arise from automated recording, the automated list is used to predict the shows that will be recorded.\(^{35}\)We assume no network guarantee on minimal views to advertisers as sometimes happens in practice. Online Appendix G considers an extension that includes rating guarantees, i.e., advertisers pay for guaranteed views instead of exposures.
\[
E_{\Theta_i} \left( \sum_t \sum_n x_{itn}^* r_{itn} \right) \geq E_{\Theta_i} \left( \sum_t \sum_n x_{itn} r_{itn} \right),
\]

where Equation (20) represents the expected cost of advertising under schedule \(\{x_{itn}^*\}\). The \(x_{itn}^*\) denote household-show selection under the optimal schedule where \(x_{itn}^* = 1\) if show \(tn\) is selected for household \(i\), \(x_{itn}^* = 0\) otherwise. Similarly, \(x_{itn} \in \{0, 1\}\) denotes household-show selection under the current schedule. \(c_{tn}\) 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 advertising viewing estimates, \(\Theta_i \equiv \{\beta_i^s, \beta_i^O, \beta_i^A, \lambda_i, c_i\}\). \(r_{itn}\) 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, and takes into account the uncertainty in advertising location within the show:

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

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

Because \(E_{\Theta_i} \left( \sum_t \sum_n x_{itn}^* r_{itn} \right) = \sum_t \sum_n x_{itn}^* E_{\Theta_i} \left( r_{itn} \right)\), we first compute \(E_{\Theta_i} \left( r_{itn} \right)\) by simulation,\(^{36}\) and then use it to solve the optimization problem for each household. The average per-view advertising price reduces from 8.0 cents to 0.7 cents. The overall cost reduces from $8.77 to $0.72, a 92% reduction in expenses. The cost reduction per 1,000 households is $9.65.

\(^{36}\)Specifically, we draw 50 sets of household-specific parameter estimates, each leading to a prediction on the household’s show choices and viewing length conditional on choice. \(E_{\Theta_i} \left( r_{itn} \right)\) is taken as the average of this predicted viewership across the 50 sets of parameter draws.
6.2.3 Profit-Based Real-time Buy

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

Incremental profits from advertising are computed using the estimates from the advertising response model in §6.1. Following Equation (18), the effect of a view of creative $a$ on household $i$’s purchase probability for brand $j$ in shopping trip $m$ is:

$$
\Delta_{ijm} = Pr\left(y^p_{ijm} = 1 \mid N^A_{ijam} = 1\right) - Pr\left(y^p_{ijm} = 1 \mid N^A_{ijam} = 0\right)
$$

$$
= \frac{\exp\left(Z_{ijm}\theta_i + h_i\left(A_{ijm} \mid N^A_{ijam} = 1\right)\right)}{1 + \sum_{j' = 1}^{J} \exp\left(Z_{ij'm}\theta_i + h_i\left(A_{ij'm} \mid N^A_{ijam} = 1\right)\right)} - \frac{\exp\left(Z_{ijm}\theta_i + h_i\left(A_{ijm} \mid N^A_{ijam} = 0\right)\right)}{1 + \sum_{j' = 1}^{J} \exp\left(Z_{ij'm}\theta_i + h_i\left(A_{ij'm} \mid N^A_{ijam} = 0\right)\right)},
$$

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

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

Figure 15 depicts the simulated profits and ROI (defined as profit per ad dollar invested) under the observed and proposed advertising schedules for different margins. The simulated short-term advertising profit using the observed schedule is negative, with losses of about $8.7 across the 834 households in our sample. In contrast, the proposed schedule produces a positive simulated incremental profit (the net gain ranges from $2.84 to $11.59). The net gain for 1,000 households ranges from $3.41 to $13.90.

---

37 Product category subscript is omitted for simplicity.

38 This lift represents the short-term effect of advertising net of purchase feedback and ad stock effects. As there is negligible difference between short- and long-term effects for the particular Charmin creative we consider (Figure 13), we abstract away from long-term effects (so this analysis provides a conservative assessment of the gains from targeting).

The total sales lift effect (short-term effect + long-term effect) is given by $\Delta_i = \Delta_{ijm} + \Delta_{ijm} \sum_{m' > m} \left[Pr\left(y^p_{ijm'} = 1 \mid y^p_{ijm} = 1\right) - Pr\left(y^p_{ijm'} = 1 \mid y^p_{ijm} = 0\right)\right]$, where the second term is the long-term effect and accounts for carry-over effect of altered choice through the purchase event feedback measure (whether the consumer purchased the brand in the previous purchase). $Pr\left(y^p_{ijm} = 1 \mid y^p_{ijm} = 1\right)$ and $Pr\left(y^p_{ijm} = 1 \mid y^p_{ijm} = 0\right)$ can be calculated using Bayes’ rule. For instance, $Pr\left(y^p_{ij,m+2} = 1 \mid y^p_{ijm} = 1\right) = Pr\left(y^p_{ij,m+2} = 1 \mid y^p_{ijm} = 1, y^p_{ij,m+1} = 1\right) Pr\left(y^p_{ij,m+1} = 1 \mid y^p_{ijm} = 1\right) + Pr\left(y^p_{ij,m+2} = 1 \mid y^p_{ijm} = 1, y^p_{ij,m+1} = 0\right) Pr\left(y^p_{ij,m+1} = 0 \mid y^p_{ijm} = 1\right)$.

Bathroom tissue has gross profit margin of roughly 13%. In the data, the mean and median prices of Charmin are respectively $6.80 and $6.30, leading to a unit dollar margin of about $0.8-0.9.
Further insights into this gain can be obtained by decomposing three factors driving the expected advertising lift, \( Pr(\gamma^A_{ln} = 1) \Delta_{jm} - c_{tn} \): 1) reducing avoidance (i.e., increasing \( Pr(\gamma^A_{ln} = 1) \)); 2) targeting people with higher advertising response (i.e., increasing \( \Delta_{jm} \)); and 3) shifting to cheaper impressions (i.e., decreasing \( c_{tn} \)). Factors 1) and 3) lower the cost of advertising, whereas factor 2) increases the revenue. To achieve this aim, we recompute profit gains under 3 alternative scenarios. In Model A, forecasted advertising avoidance is replaced by the average observed avoidance rate across viewers. In Model B, estimated advertising response is replaced by the average observed advertising response across viewers. In Model C, advertising cost is replaced by the average observed advertising cost across slots. The difference in profits between the constrained models A-C and the respective models where these effects vary across users yields the gains from targeting for each respective component (reducing avoidance, increasing ad response, and lowering costs). Figure 16 depicts this profit decomposition under different unit dollar margins. Depending on margin, the majority (75-89%) of gain comes from targeting people with higher advertising response, about 9-19% of gain comes from reducing advertising avoidance, and the remaining comes from improvement in cost efficiency. The finding that the greatest increase in profits accrues to increasing revenue by targeting those most responsive to advertising suggest the value of single-source data. Combining purchase data with viewing leads to the greatest share of profit gains.
6.2.4 Profit-Based Advance Buy

The profit-based advance buy scenario requires advertisers to forecast show viewership, advertising viewership and sales. On an intuitive level one might expect that this information requirement lowers the efficacy of advertising, because more advertisements will be aired to those who do not see them. As this statistical uncertainty reduces the expected advertising response, but not advertising expense, one might expect advertising to become less effective (thereby leading to lower optimal levels of advertising). More formally, as noted in §6.2.3, the expected incremental profit from an advertisement is $Pr (\psi^{A}_{iun} = 1) \Delta_{ijml} - c_{tn}$. In advance buy, this expected incremental profit becomes $r_{iun} \Delta_{ijml} - c_{tn}$, with $r_{iun}$ defined in Equation (23). Advertisement $tn$ is purchased for household $i$ if its expected incremental profit is positive.

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

6.2.5 Targeting Summary

The targeting results are summarized in Table 7.
Table 7: Targeting Gains

<table>
<thead>
<tr>
<th></th>
<th>Reduce Costs</th>
<th>Increase Profits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real-time</td>
<td>Target views increase 47% and costs decrease 7%; alternatively costs decrease 37% for same target views</td>
<td>Profits per thousand increase from $−10.4 to $3.4 - $13.9. Normed to a ROI of −100%, this increase implies a ROI of 108%-139%</td>
</tr>
<tr>
<td>Advance</td>
<td>Costs decrease 92% for same target views</td>
<td>Negligible spending and profit</td>
</tr>
</tbody>
</table>

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

7 Conclusion

TV remains by far the predominant form for the transmission and reception of video content, and the largest advertising medium. Moreover, its preeminence stands to benefit from recent digital innovations such as DVRs and STBs. Yet digital TV is both a blessing and a curse for advertisers. On the one hand, set-top boxes have greatly enhanced viewers’ TV-viewing experiences, leading to increased viewing consumption and the opportunity to target advertising at the TV set level. On the other hand, these boxes enable viewers 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 set-top boxes 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 (set-top box 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 views 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 raise incremental profit even in the face of ad avoidance. We find that the greatest potential to increase the profitability of advertising arises from i) the integration of purchase data with viewing data and ii) the ability to buy placements real-time instead of in advance. As for the former, single source data are becoming increasingly available and purveyed by firms such as Axciom TV. As for the latter, there is a developing advertising ecosystem that could enable real-time buying as is commonly observed in display advertising (wywy (2016)). Our analysis suggests the importance of these technological advances.

We note that our data were collected almost 10 years ago, when the field was in its infancy. To our knowledge, ours was the only single source data in existence at that time and there were no tools for targeting. Now the data are ubiquitous and the technology has become available, so our approach is more actionable than ever.

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. (2016), the quantity of the advertised product purchased recently can explain the advertising avoidance 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.

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.

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

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. 2016. Us digital ad spending to surpass tv this year.


wywy. 2016. Programmatic tv: How it works, the players & the right strategies. *wywy*.


Yao, Song, Wenbo Wang, Yuxin Chen. Forthcoming. Tv channel search and commercial breaks. *Journal of Marketing Research*.
Online Technical Appendix

A Heterogeneity in Show Uncertainty

We demonstrate that our show utility specification captures both heterogeneity in both individual priors and show priors.

**Individual Priors** Our approach captures viewer-specific heterogeneity in familiarity non-parametrically by estimating the model at the person-specific level, implying the variance terms of $\varepsilon_{itn}$ are free to vary by person, $i$, though this variance is not separately identified in the logit from the preference parameters. More formally, consider writing show utility at the viewer level as:

$$u_{itn} = X_{itn}\beta_i + \nu_{itn} + \varepsilon_{itn}$$

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

where show $n$ is sampled if $\bar{u}_{itn} + \nu_{itn} > \bar{u}_{itn}' + \nu_{itn}'$, $\forall n'$. Assuming $\text{Var}(\nu_{itn}) = \sigma_i^2$, such that there is a person specific variance to the show implies a sampling probability of

$$Pr(y_{itn} = 1) = \frac{\exp\left(\bar{u}_{itn}' / \sigma_i\right)}{\sum_{n'} \exp\left(\bar{u}_{itn}'' / \sigma_i\right)}$$

$$= \frac{\exp\left(X_{itn} \beta_i / \sigma_i\right)}{\sum_{n'} \exp\left(X_{itn} \beta_i / \sigma_i\right)}$$

That is, the ratio $\beta_i / \sigma_i$ is estimated. Though $\beta_i$ and $\sigma_i$ are not separately identified, the model accommodates heterogeneous uncertainties non-parametrically. People with better knowledge have smaller $\sigma_i$ and larger

$$\beta_i^* = \frac{\beta_i}{\sigma_i}.$$

**Show Priors** Denote show series loyalty/experience/familiarity as $Z_{itn}$, which measures the number of previous episodes of the program that the household has sampled in the preceding week. It might be
expected that uncertainty decreases with familiarity. Hence, we assume \( \nu_{itn}^* = \frac{\nu_{itn}}{Z_{itn}} \), so that \( Var(\nu_{itn}^*) = \left( \frac{\sigma_i}{Z_{itn}} \right)^2 \). Rewrite show utility as follows:

\[
    u_{itn} = X_{itn} \beta_i + \nu_{itn}^* + \varepsilon_{itn}^*.
\]

Then \( Var(Z_{in} \nu_{in}) = \sigma_i^2 \), and the sampling probability can in turn be derived as:

\[
    Pr(y_{itn} = 1) = Pr \left\{ X_{itn} \beta_i + \nu_{itn}^* > X_{itn'} \beta_i + \nu_{itn'}^*, \forall n' \right\}
    = Pr \left\{ X_{itn} Z_{itn} \beta_i + Z_{itn} \nu_{itn}^* > X_{itn'} Z_{itn} \beta_i + Z_{itn} \nu_{itn'}^*, \forall n' \right\}
    = \frac{\exp \left( X_{itn} Z_{itn} \beta_i / \sigma_i \right)}{\sum_{n'} \exp \left( X_{itn'} Z_{itn'} \beta_i / \sigma_i \right)}.
\]

This is equivalent to results based on the following utility specification:

\[
    u_{itn} = X_{itn} Z_{itn} \beta_i + \nu_{itn} + \varepsilon_{itn}.
\]

Hence by adding interaction terms between loyalty and other covariates, we essentially allow for heterogeneity in show uncertainty.

Then the familiarity experience adjusted error variance term, \( (\sigma_{itn}^* = \sigma_i / Z_{itn}) \) changes with familiarity. Specifically, as the shows becomes more familiar the uncertainty decreases, \( Z_{itn} \uparrow \Rightarrow \sigma_{itn}^* = \frac{\sigma_i}{Z_{itn}} \downarrow \). Moreover, because the model is estimated at the person level, this also implies a show-person interaction.

### B 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 $t$, $\frac{t_m}{t}$

Then:

$$Pr\{l \geq t\} = Pr\{N_{exit}(t) = 0\}$$

$$= \sum_{k=0}^{\infty} Pr\{N(t) = k\} \sum_{i=1}^{k} Pr\{N_1(t) = n_1, N_2(t) = n_2, \ldots, N_M(t) = n_M\} \times \prod_{m=1}^{M} (p_m)^{n_m}$$

$$= \sum_{k=0}^{\infty} e^{-\lambda t} (\lambda t)^k \frac{k!}{k!} \prod_{m=1}^{M} \left[ q_m(t) p_m \right]^{n_m}$$

$$= \sum_{k=0}^{\infty} e^{-\lambda t} (\lambda t)^k \frac{k!}{k!} \prod_{m=1}^{M} \left[ q_m(t) p_m \right]^{n_m}$$

$$= e^{-\lambda t} \sum_{k=0}^{\infty} \left\{ \frac{\lambda t}{\sum_{m=1}^{M} q_m(t) p_m} \right\}^k$$

$$= e^{-\lambda t} e^{\lambda t [\sum_{m=1}^{M} q_m(t) p_m]}$$

$$= e^{-\lambda t [\sum_{m=1}^{M} \frac{t_m}{t} p_m]}$$

$$= e^{-\lambda \sum_{m=1}^{M} t_m [1-p_m]}$$

C Viewing Model Identification and Likelihood Function

C.1 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 ($\beta^A_i$) 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 $u^S_{itn}$ for all $i, t, n$). For the second, the opportunity cost of time is too high (i.e., large $u^O_{it}$). 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 ($\beta^S_i$) 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.

We acknowledge the caveat that identification is subject to various parametric assumptions: $\nu^S_{itn}$ is i.i.d. standard Type I Extreme Value distributed; $\epsilon^S_{itn}$ is i.i.d. Type I Extreme Value distributed with mean zero; $\epsilon^A_{itn}$ is i.i.d. standard Type I Extreme Value distributed; $\nu^O_i$ is i.i.d. standard Type I Extreme Value distributed and is i.i.d. with $\nu^S_{itn}$; arrivals of external shocks follow a homogeneous Poisson process with rate $\lambda_{itn} = \exp(g_{tn}\rho_i)$. 

Marketing Science Institute Working Paper Series
C.2 Likelihood Function

We estimate the viewing model using Matlab by simulated maximum likelihood approach. The sampling order implies $v_{itn}^S$ and $v_{it}^O$ in Equations (5), (8), and (13) are truncated respectively below $\bar{u}_{itn}^S + v_{itn}^S - \bar{u}_{itn}'$ and below $\bar{u}_{itn} + v_{itn}^S - \bar{u}_{it}^O$. In the estimation, we first simulate $K = 100$ sets (Chen and Yao (Forthcoming)) of $(v_{itn}^S, v_{it}^O)$, denoted as $\{v_{itn}^{S1}\}_{n}, \ldots, \{v_{itn}^{SK}\}_{n}, v_{it}^{O1}, \ldots, v_{it}^{OK}$. 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 (i.e., viewed for 30 seconds or more) contributes to the likelihood with:

$$Pr\left(y_{itn}^S = 1\right) = \frac{\exp\left(X_{itn}^S \beta_i^S\right)}{\exp\left(X_{it}^O \beta_i^O\right) + \sum_{n' \in N_i} \exp\left(X_{itn'}^S \beta_i^S\right)}.$$

Watching

Each sampled show that is watched (i.e., viewed for 3 minutes or more) contributes to the simulated likelihood with:

$$Pr\left(y_{itn}^W = 1 \mid y_{itn}^S = 1\right) = Pr\left(u_{itn}^S \geq IV_{itn} \epsilon_i\right)$$

$$= \frac{1}{K} \sum_{k=1}^{K} \left[ 1 - F_{u_{itn}^S}^\epsilon \left( \max \left\{ \bar{u}_{itn}^O + v_{itn}^{Ok}, \ln \left( \sum_{n' \in N_i} \exp\left(\bar{u}_{itn'}^S + v_{itn'}^S\right)\right) \right\} - \bar{u}_{itn}^S - v_{itn}^S \right],$$

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:

$$Pr\left(l_{itn}^W \mid l_{itn}^W = L_{itn}\right) = Pr\left\{l_{itn}^* > L_{itn}\right\}$$

$$= \frac{1}{K} \sum_{k=1}^{K} e^{-\lambda_{itn} \sum_{m=1}^{L_{itn}} \left( q_{itn}^m \mid \{v_{itn}^{Sk}\}_{n'}, \bar{u}_{itn}' \right)}.$$
where:

\[ q_{itn}^m \mid \{ \nu_{itn'} \}_{n'} \}, \nu_{it}^O \]

\[ = F_{\nu_{it}^M} \left( \max \left\{ \tilde{u}_{it}^O + \nu_{it}^O, \ln \left( \sum_{n' \in N_{it}} \exp \left( \tilde{u}_{itn'}^O + \nu_{itn'}^O \right) \right) \right\} - \tilde{u}_{itn}^O - \nu_{itn}^O \right) . \]

Each show that is not watched until the end \((l_{itn}^W < L_{itn})\) contributes to the likelihood with:

\[ Pr \left( l_{itn}^W \mid l_{itn}^W < L_{itn} \right) = f \left( l_{itn}^W \right) = \frac{1}{K} \sum_{k=1}^{K} \lambda_{itn}^{\tilde{m}} q_{itn}^{\tilde{m}} e^{-\lambda_{itn} \sum_{m=1}^{M} \nu_{itn}^m \left( q_{itn}^m \mid \{ \nu_{itn'}^m \}_{n'}, \nu_{it}^O \right)} , \]

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

**Recording**

Each recorded show contributes to the likelihood with:

\[ Pr \left( y_{itn}^R = 1 \right) = \frac{\exp \left( X_{itn}^S \beta_i^S \right)}{\sum_{n'} \exp \left( X_{itn'}^S \beta_i^S \right)} . \]

**Advertising zapping (live shows)**

Each non-zapped advertisement contributes to the likelihood with:

\[ Pr \left( y_{itn}^{AL} = 1 \right) = 1 - \frac{1}{K} \sum_{k=1}^{K} F_{\nu_{itn}^A} \left( \max \left\{ \tilde{u}_{it}^O + \nu_{it}^O, \ln \left( \sum_{n' \in N_{it}} \exp \left( \tilde{u}_{itn'}^O + \nu_{itn'}^O \right) \right) \right\} \\
- c_i - \tilde{u}_{itn}^A \right) \right) , \]

and each zapped advertisement contributes to the likelihood with:

\[ Pr \left( y_{itn}^{AL} = 0 \right) = 1 - Pr \left( y_{itn}^{AL} = 1 \right) . \]

**Advertising zipping (recorded shows)**

Each non-zipped advertisement contributes to the likelihood with:

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

and each zipped advertisement contributes to the likelihood with:
\[ Pr\left( y_{itn}^R = 0 \right) = 1 - Pr\left( y_{itn}^R = 1 \right). \]

Taken together, the likelihood function of \((\beta_i, \beta_i^C, \eta_i, \rho_i, \beta_i^A, c_i)\) can be written as follows:

\[
L \left( \beta_i, \beta_i^C, \eta_i, \rho_i, \beta_i^A, c_i \mid \{ y_{itn}^S, y_{itn}^W, l_{itn}^W, y_{itn}^R, y_{itn}^{AL}, y_{itn}^{AR} \} \right) \\
= \prod_{t,n} \left\{ Pr\left( y_{itn}^S = 1 \right) \right\}^{y_{itn}^S} \times \left[ Pr\left( y_{itn}^W = 1 \mid y_{itn}^S = 1 \right) \right]^{y_{itn}^W = y_{itn}^S = 1} \times \left[ Pr\left( y_{itn}^W = 0 \mid y_{itn}^S = 1 \right) \right]^{y_{itn}^W = y_{itn}^S = 0} \times \left[ Pr\left( y_{itn}^W = 0 \mid l_{itn}^W < L_{itn} \right) \right]^{l_{itn}^W < L_{itn}} \times \left[ Pr\left( y_{itn}^W = 0 \mid l_{itn}^W = L_{itn} \right) \right]^{l_{itn}^W = L_{itn}} \times \left[ Pr\left( y_{itn}^R = 1 \right) \right]^{y_{itn}^R} \times \left[ Pr\left( y_{itn}^{AL} = 1 \right) \right]^{y_{itn}^{AL}} \times \left[ Pr\left( y_{itn}^{AL} = 0 \right) \right]^{1 - y_{itn}^{AL}} \times \left[ Pr\left( y_{itn}^{AR} = 1 \right) \right]^{y_{itn}^{AR}} \times \left[ Pr\left( y_{itn}^{AR} = 0 \right) \right]^{1 - y_{itn}^{AR}}.
\]

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

### C.3 Simulation

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>Table A.1: Monte Carlo Simulation Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
</tr>
<tr>
<td>(\beta_i^S)</td>
</tr>
<tr>
<td>(\beta_i^S)</td>
</tr>
<tr>
<td>(\beta_i^C)</td>
</tr>
<tr>
<td>(\beta_i^A)</td>
</tr>
<tr>
<td>(\lambda)</td>
</tr>
<tr>
<td>(c)</td>
</tr>
</tbody>
</table>
D 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.

![Empirical CDF (Across Households) of the Hit Rate on Show Sampling (a) and Watching (b)](image)

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.


E  Purchase Data Summary and a Model-free Sales Response Model

E.1  Purchase Data Summary

To measure advertising response, we select seven product categories that evidence high variation in advertising and sales so as to enhance the likelihood of measuring advertising effects: 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. Table A.2 presents descriptive information on these categories.

E.2  Model Free Analysis

We apply a “model-free” approach to measure advertising effects. We start with homogeneous advertising effects, then allow advertising effects to vary across consumer segments.
E.2.1 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 (e.g., Tellis (1988); Mela et al. (1997)). Moreover, there are problems of advertising attribution, wherein 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. The downside of these approaches is that they involve numerous assumptions regarding functional form and error distributions. In contrast, our rich individual level panel data enable a different tact.

Specifically, we can hold the in-store causal environment constant by considering two same-store visits within the same week. In the cases where a household did not purchase on the first trip, but did on the second, the only major factors that can vary between trips are household inventory and advertising. 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 an advertisement 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, pooling household observations for each brand, we compute each brand’s second-visit purchase log odds ratio with and without advertising views received between the two visits (Figure A.3). 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.

---

40 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 previous visit.

41 Let $N_a$ indicate the number of people exposed to a brand’s advertising between the first and second visit, of which $n_a$ purchased in the second visit. Similarly, denote $N_n$ as the number of people who were not exposed to advertising between the first and second visit, of which $n_n$ purchased in the second visit. Then the log odds ratio of the fraction who purchased with ad exposure over the fraction who purchased without ad exposure is $\ln \left( \frac{n_a/N_a}{n_n/N_n} \right)$. If there is no advertising effect, then in expectation, $\frac{n_a}{N_a} = \frac{n_n}{N_n}$ and $\ln \left( \frac{n_a/N_a}{n_n/N_n} \right) = 0$.

Marketing Science Institute Working Paper Series
for most brands. A one-sample Kolmogorov-Smirnov test indicates the brand-level log odds ratio is not normally distributed with mean zero (p<0.016).

![Histogram of Purchase Log Odds Ratio With and Without Advertising Views (by Brand)](image)

Figure A.3: Histogram of Purchase Log Odds Ratio With and Without Advertising Views (by Brand)

As a robustness check, we consider endogeneity bias that might arise from targeting households that are more responsive to advertising (e.g., Narayanan and Manchanda (2009)). Exploiting the panel nature of our data, we estimate a linear probability model for each brand:

\[
y_{ijm}^P = \alpha_{ij} + \gamma_j A_{ijm} + \epsilon_{ijm},
\]

where \(i\) denotes household, \(j\) denotes brand, and \(m\) 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 \(i\) is exposed to brand \(j\)'s advertisement since the preceding trip (in the same store and week). \(\alpha_{ij}\) is a household fixed effect that captures unobserved heterogeneity, and \(\epsilon_{ijm}\) is an idiosyncratic error term. In this model, the parameter \(\gamma_j\) measures the advertising lift. A two-sample Kolmogorov-Smirnov test indicates the distribution of the estimated advertising lift across brands \((1, ..., J)\) \(\hat{\gamma}_j\), is not significantly different from the distribution of \(\hat{\gamma}_j\) obtained under an alternative model without household fixed effects. This suggests results are invariant to the inclusion of household fixed effects that control for unobserved household-specific factors (e.g., demographic based targeting).

Our second robustness check addresses the concern that the gap between the two visits might drive both advertising view and purchase likelihood in the same direction. 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. To see if this is the case, we compute purchase log odds ratio by brand and gap between visits (1 or 2 days, 3 or 4 days, and 5-6 days). If the identified advertising effect were simply correlation induced by the gap between two visits, then within each gap group, we would expect the distribution of log odds ratio across brands to be centered around zero. Instead, we find the distribution within each gap group to be skewed to the right of zero. Therefore, the causal effect of advertising is valid.

E.2.2 Heterogeneous Advertising Effects

In the preceding subsection, we presumed that the increase in a brand’s purchase probability conditioned on an advertising view is identical across households. We extend that model to consider a latent class representation of heterogeneity in advertising response. For household \(i\) who belongs to segment \(k\) \((k = 1, ..., K)\), the second-visit probability of purchasing brand \(j\) is specified as:

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

(E.1)

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 \(i\) is exposed to 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:

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

(E.2)

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:

\[
Pr\left( i \in k \mid \{y_{ijm}^P\}_{j,m} \right) = \frac{Pr\left(i \in k\right) \prod_m \prod_j \left(Pr\left(y_{ijm}^P = 1 \mid i \in k\right)\right)^{y_{ijm}^P}}{\sum_{k'} Pr\left(i \in k'\right) \prod_m \prod_j \left(Pr\left(y_{ijm}^P = 1 \mid i \in k'\right)\right)^{y_{ijm}^P}}.
\]

(E.3)

The likelihood function associated with household \(i\)’s purchase decisions is:
\[ L_i = \sum_k Pr(i \in k) \prod_m \prod_j (Pr(y_{ijm}^p = 1 | i \in k))^y_{ijm}, \tag{E.4} \]

and the overall log likelihood is:

\[ \ln L = \sum_i \ln L_i. \tag{E.5} \]

Because different product categories are associated with different purchase frequencies (Table A.2), 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 rest two product categories (regular cola and diet cola), three segments are identified. Table A.3 reports the estimated advertising effect (\(\theta_k\)) by product category and consumer segment.

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

**F An Aggregate Level Targeting Model**

Current targeting practice consists of *advance* media buys by the show’s demographic and anticipated viewing level. Corresponding to this practice, we consider an aggregate-level targeting model that maximizes advertising views given a cost where advertisers can only target shows.

The advertiser selects shows to advertise in to maximize total exposures for sample households. As before, a show is represented by time (\(t\)) and network (\(n\)). \(x_{tn}\) denotes show selection. \(x_{tn} = 1\) if show \(tn\) is selected, \(x_{tn} = 0\) otherwise. Given the budget \(B\) (which equals current advertising spending), the optimization problem can be written as:

\[ \text{Max}_{\{x_{tn}\}}, n \sum_i \sum_n x_{tn}r_{itn} \]

such that

\[ x_{tn} \in \{0, 1\}, \forall t, n \]

\[ \text{(F.1)} \]

\[ \text{(F.2)} \]
where \( c_{tn} \) is the advertising price associated with show \( tn \), and \( r_{itn} \) is the probability that household \( i \) watches the advertisement placed in show \( tn \), and can be obtained from the viewing model output:

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

where \( t' \) denotes a possible advertising location (time into show) and \( f(t') \) represents the probability density function of a uniform distribution, whose support is advertising pods in show \( tn \). This distribution is used to capture advertiser uncertainty regarding where in the pod its advertisement will appear, as this position is unknown to advertisers at the time of purchase. 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 \( t' \). \( Pr \left( y_{itn}^A = 1 \right) \) denotes the probability that this advertisement will not be zapped (if the show is live) or zipped (if the show is recorded). Note that Equation (F.3) is the budget constraint and Equation (F.2) ensures binary assignment for advertisements into slots. Equation (F.1) indicates the total aggregate revenues of an advertising schedule given by \( x_{tn} \).

We implement this targeting strategy for Charmin in the hold-out period of July 2006. We account for uncertainty in viewing model estimation with Monte Carlo simulation, drawing 50 sets of household-specific parameter estimates. Each parameter set leads to a prediction on households’ show choices and viewing length conditional on show choice. Aggregating across households results in predicted viewership of different shows (where ads are nested in shows). This predicted viewership, along with an advertising price index for each show, is used to solve this assignment problem given by Equations (F.1) to (F.3). This results in 50 sets of optimal advertising schedules, each corresponding to a Monte Carlo draw of a parameter set.

We solve the optimization problem is solved respectively for two sets of households: the entire set of 834 panelists, and the 122 households observed to behave as non-loyals in the bathroom tissue category. In order to evaluate the performance of optimal schedules, we compare the kernel densities (across the 50 sets of parameter estimate draws) of predicted and observed Charmin advertising views under current and

\[\text{The advertising price index for a show is the median observed price for advertisements in the show.}\]
optimal advertising schedules, shown in Figure A.4. “Predicted” views are computed based on predicted viewership from Equation (F.4), and “observed” views are obtained from observed viewership in the data. Measured by observed views, the optimal schedules significantly outperform the current schedule for both sets of households. Note that the model tends to overpredict exposures, though not significantly so. That said, the improvement in exposures arising from the optimal schedule far exceeds this small prediction error and therefore these gains are significant.

Figure A.4: Kernel Densities of Predicted and Observed Charmin Advertising Views under Current and Optimal Advertising Schedules. Panel (a) is based on optimizations over all panelists, and Panel (b) is based on optimizations over non-loyals. The distributions are across the Monte Carlo draws of viewing model estimates.

To further explore differences between optimal and current schedules, Figures A.5 and A.6 respectively compare the network shares and genre shares in optimal and current schedules. In terms of network selection, optimal schedules feature a much higher share of NBC. In terms of genre selection, optimal schedules feature higher shares of dramas, news and sports, and lower shares of comedies and reality TV shows.

Notably, optimal schedules differ between all households and non-loyals. For instance, Figure A.5 indicates a lower share of ABC and a higher share of NBC for non-loyals. The two sets of schedules share seven of the top ten programs (pooled across optimal schedules under all sets of parameter draws) and differ in the other three. These differences suggest people with dissimilar brand preference might prefer different shows, which can be leveraged in targeting through advance buy.
Figure A.5: Boxplots of Network Shares in Optimal and Current Advertising Schedules. Each boxplot depicts the distribution of the corresponding network’s share in optimal schedules. For each network, the two boxplots are based respectively on optimizations over all panelists and over non-loyals. The dashed segments represent network shares in current schedule.

G Targeting Under Rating Guarantees

We consider targeting models under the alternative assumption that advertisers do not pay for avoided advertisements (i.e. viewing guarantees). We then explore the extent to which advertising prices can increase while keeping advertisers indifferent between this new policy and one wherein advertisers do have to pay for avoided ads. With viewing guarantees, the viewing model is no longer relevant as there is no cost for a non-view meaning the implied viewing probability is one. In addition, because viewership uncertainty no longer exists, cost-based real-time buy and cost-based advance buy result in the same allocation and number of views. However, profit-based real-time buy and profit-based advance buy are not necessarily the same because in the former case, we have information on the household’s realized purchase behavior up to the advertising slot.

G.1 Cost-Based Approach

In this targeting scenario, the advertiser will simply select shows that have the lowest per-view cost. An alternative schedule that yields identical views for each household within the same month can lower CPM from $12.7 to $6.14, a 51% reduction in expenses. Accordingly, TV networks can raise per-view price by 51% to make the advertiser indifferent from the current practice.
If the advertiser seeks to maintain the number of daily views, CPM can still be decreased to $10.49, or by 18%. Accordingly, TV networks can raise per-view price by 18% to make the advertiser indifferent from the current practice.

G.2 Profit-Based Real-time Buy

This targeting scenario is similar to §6.2.3, except that the viewership probability affects both expected revenue and expected cost. Without viewing guarantees, the expected incremental profit is $Pr(y_{itn}^A = 1) \Delta_{ijn}l - c_{tn} = Pr(y_{itn}^A = 1) \Delta_{ijn}l - r_{itn}c_{tn} - (1 - r_{itn})c_{tn}$, meaning the advertiser pays for the advertisement no matter it is viewed (with probability $Pr(y_{itn}^A = 1)$) or not (with probability $1 - Pr(y_{itn}^A = 1)$). Because of viewing guarantees, advertisements that are not viewed will not be paid for; that is, the cost is 0 with probability $1 - Pr(y_{itn}^A = 1)$ and the latter term becomes $(1 - Pr(y_{itn}^A = 1))c_{tn} = (1 - Pr(y_{itn}^A = 1))0 = 0$. Therefore, the expected incremental profit becomes $Pr(y_{itn}^A = 1) \Delta_{ijn}l - Pr(y_{itn}^A = 1)c_{tn}$. As a result of decreased cost, more ad slots will be profitable.

Figure A.7 depicts the simulated profits and ROI (defined as dollar of profit per dollar invested) 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 of about $8.7 across the 834 households in our sample. The proposed schedule produces a positive simulated incremental profit.
ranging from $5.1 to $18.7, and the profit per thousand households ranges from $6.1 to $22.4. If TV networks were able to change price under this ad-buy model, then they can raise price to extract all surplus from the advertiser. The average per-view price is therefore obtained by dividing the profit difference between the optimal schedule and the current schedule with the number of ads under the optimal schedule, and would be 44-68% higher than the current level.

![Graph showing simulated incremental profit and ROI under different unit dollar margins.]

Figure A.7: Simulated Incremental Profit and ROI (for Sample Households, Real-time Buy) Under Different Unit Dollar Margins

**G.3 Profit-Based Advance Buy**

In the profit-based advance buy scenario, a similar approach is adopted as in the profit-based real-time buy scenario. Similarly, the expected incremental profit changes from $r_{itn}\Delta_{ijml} - c_{tn}$ when there is no viewing guarantees to $r_{itn}\Delta_{ijml} - r_{itn}c_{tn}$. As a result of decreased cost, more ad slots become profitable. We find the proposed optimal schedule produces a small, but positive expected incremental profit (The profit per thousand households ranges from $0.08 to $0.40). The reason why the expected profit is much lower than in the profit-based real time buy is because the expected viewing probability is low when planning in advance, leading to low expected incremental profit. If TV networks were able to change price under this ad-buy model, then they can raise price to extract all surplus from the advertiser. The average per-view price is therefore obtained by dividing the profit difference between the optimal
schedule and the current schedule with the number of ads under the optimal schedule, and would be 124-343% higher than the current level.
Table A.2: Purchase Data Summary

<table>
<thead>
<tr>
<th>Product Category</th>
<th>Number of Households</th>
<th>Number of Shopping Trips</th>
<th>Number of Purchases</th>
<th>Brands in Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Children’s yogurt</td>
<td>120</td>
<td>9,827</td>
<td>402</td>
<td>Yoplait Trix, Dannon Danimals, Yoplait Go Gurt</td>
</tr>
<tr>
<td>Children’s cereal</td>
<td>525</td>
<td>47,022</td>
<td>2,829</td>
<td>Quaker Cap’n Crunch, Kellogg’s Froot Loops, Kellogg’s Frosted Flakes, General Mills Lucky Charms, General Mills Cinnamon Toast Crunch</td>
</tr>
<tr>
<td>Regular cola</td>
<td>446</td>
<td>40,679</td>
<td>3,189</td>
<td>Coke, Pepsi</td>
</tr>
<tr>
<td>Diet cola</td>
<td>416</td>
<td>38,409</td>
<td>4,071</td>
<td>Coke, Pepsi</td>
</tr>
<tr>
<td>Sports drink</td>
<td>298</td>
<td>26,874</td>
<td>1,377</td>
<td>Gatorade, Powerade</td>
</tr>
<tr>
<td>Toothpaste</td>
<td>386</td>
<td>39,268</td>
<td>1,017</td>
<td>Crest, Colgate, Aquafresh</td>
</tr>
<tr>
<td>Bathroom tissue</td>
<td>599</td>
<td>53,738</td>
<td>2,949</td>
<td>Charmin, Angel soft, Cottenelle, Quilted Northern, Scott</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Product Category</th>
<th>Number of Purchases Per Household</th>
<th>Fraction of Purchases Made at Regular Price</th>
<th>Total Number of Weekly Advertisements</th>
<th>Average Number of Weekly Advertising Views Per Household</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
</tr>
<tr>
<td>Children’s yogurt</td>
<td>3.4</td>
<td>0.84</td>
<td>0.37</td>
<td>104.3</td>
</tr>
<tr>
<td>Children’s cereal</td>
<td>5.4</td>
<td>0.50</td>
<td>0.50</td>
<td>152.8</td>
</tr>
<tr>
<td>Regular cola</td>
<td>7.2</td>
<td>0.21</td>
<td>0.41</td>
<td>155.7</td>
</tr>
<tr>
<td>Diet cola</td>
<td>9.8</td>
<td>0.24</td>
<td>0.43</td>
<td>150.8</td>
</tr>
<tr>
<td>Sports drink</td>
<td>4.6</td>
<td>0.37</td>
<td>0.48</td>
<td>167.1</td>
</tr>
<tr>
<td>Toothpaste</td>
<td>2.6</td>
<td>0.70</td>
<td>0.46</td>
<td>870.9</td>
</tr>
<tr>
<td>Bathroom tissue</td>
<td>4.9</td>
<td>0.58</td>
<td>0.49</td>
<td>508.4</td>
</tr>
</tbody>
</table>

Table A.3: Advertising Effect by Product Category and Consumer Segment

<table>
<thead>
<tr>
<th>Product Category</th>
<th>Segment 1 Mean (S.D.)</th>
<th>Segment 2 Mean (S.D.)</th>
<th>Segment 3 Mean (S.D.)</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.