Complementarities in Consumption and the Consumer Demand for Advertising*

Anna E. Tuchman
Kellogg School of Management

Harikesh S. Nair
Stanford GSB

Pedro M. Gardete
Stanford GSB

Previous Versions: Oct 2014; June 2015; This version: Oct 10, 2016†

Abstract

Leveraging new data that links TV ad-viewing with product purchases for a large sample of households, we provide empirical evidence that is consistent with Becker and Murphy's (1993) proposal that advertising can act as a complement to product consumption. Under an exclusion restriction that product prices do not affect the utility from consuming advertising directly, we document that ad-skipping is lower when a household has purchased more of the advertised brand. Fitting a structural model that incorporates such complementarities to the data, we evaluate consumer welfare in advertising targeting counterfactuals motivated by the "addressable" future of TV. Reflecting the positive view of advertising in the model, we find that net consumer welfare may increase in several scenarios. This occurs because under improved targeting, firms shift advertising to those who are likely to value it. At the same time, consumers that do not value the ads end up skipping them, mitigating possible welfare losses. Both forces are relevant to assessing advertising effects in a world with improved targeting and ad-skipping technology.

Keywords: Advertising, ad-skipping, addressable TV, complementarities, treatment effects, non-compliance, discrete-continuous demand, consumer welfare.

---

*We thank Magid Abraham, Kyle Bagwell, Lanier Benkard, J-P. Dubé, Gautam Gowrisankaran, Günter Hitsch, Kirthi Kalyanam, Sanjog Misra, Martin Pietz, Rahul Telang, Mo Xiao, Song Yao, David Zvilichovsky; seminar participants at Chicago-Booth, Duke-Fuqua, Erasmus, Northwestern-Kellogg, Insead, Stanford-GSB, Temple-Fox, UCLA-Anderson, Michigan-Ross, Univ. of Arizona, and Univ. of Washington-Foster; participants at the 2015 SICS (Berkeley), 2015 NBER-IO (Stanford), the 2014 QME (USC), Marketing Science (Atlanta), TADC (LBS), and Economics of ICT (Mannheim) conferences; and especially Peter Rossi and Joel Waldfogel for useful comments. We thank the Wharton Customer Analytics Initiative and an anonymous sponsoring firm for generously making the data available for academic research. The usual disclaimer applies. Authors are listed in reverse alphabetical order and contributed equally. Please contact Tuchman (`anna.tuchman@kellogg.northwestern.edu`), Nair (`harikesh.nair@stanford.edu`) or Gardete (`gardete@stanford.edu`) for correspondence.

†This version reports new instrumental variables specifications and details on price variation that pertain to identification, based on information we obtained recently on the institutional context of our data.

"Advertising has always been a difficult subject to introduce into the conventional theory of consumer choice." − Auld (1974), *The Quarterly Journal of Economics.*

1 Introduction

Markets for advertising now make up a large part of the economy. Advertising revenues in the United States for 2015 totaled more than $181 billion, with Internet advertising totaling $60 billion and TV advertising (Broadcast and Cable TV combined) totaling $66 billion in revenues (IAB 2016). Many modern markets for online search and social networking, broadcast TV, magazine and print media are sustained by advertising revenues. Against this background, the study of advertising and how it affects behavior is now one of the key problems of interest to firms and one of the important questions of Economics and Marketing.

While questions of how to target, measure and determine mechanisms to sell advertising have been studied by academics, the question of how recipients of ads choose to consume advertising has received scarce attention. Although exposure to advertising is at least partially under the control of firms, the *consumption* of advertising is ultimately under the control of consumers. Consumers can discard direct mail they do not value, skip TV ads they do not enjoy, or scroll away from online video ads they find a nuisance. Surprisingly, however, most of the empirical literature has ignored the role of consumer choice over ad consumption, treating the level of advertising an agent sees as determined primarily by the sophistication of firms' targeting technology and the *supply* of advertising, ignoring consumer *demand* for the ads. Viewing ad consumption as a choice by consumers changes the way we assess the effects of advertising, the mechanisms by which advertising works, as well as the assessment of the welfare effects of advertising. This paper presents new data on TV advertising consumption to show evidence for ad choice by consumers, develops a model for the choice of advertising consumption based on Becker and Murphy's (1993) theory of complementarities for which we find support, and presents estimates from the model to illustrate the role of ad choice in assessing advertising effects and welfare.

A formal treatment of advertising consumption requires a precise model of the decision problem a consumer solves in order to determine whether or not to watch an ad. The framework developed by Becker and Murphy has intuitive appeal for this reason. The Becker-Murphy approach is to treat advertising as a good in consumers' utility functions, thereby creating a role for consumer choice over advertising consumption. In this framework, advertising affects product demand due to complementarities in the joint consumption of advertising and products, rather than by shifting consumers' tastes or providing information that changes consumers' beliefs.[1] Complementarities imply that consuming more of the

[1]Past frameworks for handling the micro-foundations of advertising include the informative model that posits that advertising affects demand by communicating information about products to consumers (Nelson 1970; 1974; Butters 1977; Grossman and Shapiro 1984), and the so-called persuasive model, in which advertising is incorporated into the utility from *product* consumption and viewed as a means of creating brand loyalty (please see Bagwell 2007 for a comprehensive

advertised product increases the marginal utility from ad consumption, so observed purchase quantities are informative in explaining advertising consumption. Thus, utilizing the theory helps leverage the predictive power of purchase data in an internally consistent way to understand ad choice. The link to a well-defined utility maximization problem also makes a precise characterization of the consumer demand for advertising possible.

Viewing the advertising choice problem this way has three main implications for empirical analysis. First, modeling advertising as a choice changes how we assess the welfare effects of ads. Advertising in this framework is an endogenous avoidable choice. Individuals consume ads only if it increases their net individual welfare. Consumers who do not obtain value from consuming ads would avoid them upon initial exposure. The typical treatment of advertising assumes ad consumption is the same as ad exposure (because the choice to consume advertising is not modeled). In our set-up, ad exposures and ad consumption are separate constructs. Compared to the typical set-up, our model presents a more positive role for advertisements because active avoidance of annoying ads reduces the potential for welfare losses, and consumption of advertisements positively affects welfare by inducing higher product demand and consumption. In a world with improved ad-avoidance technology and significant ad-skipping, this positive perspective may become more relevant than ever before.

Second, the fact that advertising consumption is actively chosen by consumers complicates the assessment of the causal effects of advertising. Even if ad exposure is randomized, the treatment − ad consumption − is not, because those that are more likely to like the ads end up seeing them. This is a problem of non-compliance. A precise model of who takes up treatment is then useful to assess the full distribution of treatment effects, as well as to characterize the sub-populations to which the measured treatment effects apply.

Finally, as we explain in the next section, a formal assessment of who would consume ads and the extent to which ad consumption shifts purchase behavior is important for the individual-level targeting of advertising. Such targeting is becoming common in TV ad-markets. As TV becomes more "addressable" (for example, via set-top boxes and internet IP-address enabled viewing devices), TV ad markets now allow advertisers to target advertisements to specific consumers based on their observed historical product purchase and ad viewing behavior (see Perlberg 2014; O'Connor 2014). Compared to the traditional demographic categories of age and gender, companies like *Nielsen Catalina Solutions* now merge credit card data from shopper loyalty cards with TV viewership data and provide advertisers and networks a behavior-based profile of what kind of viewer is buying each type of consumer packaged good. DirectTV

review of the literature). The informative view is not a good description of ad consumption in our study. The product category we study is a fast moving consumer packaged good that has been on the market for years with no new brand entry during the time period of our data. Like Ackerberg (2001), we find that advertising continues to affect the purchase behavior of experienced consumers in the data even after significant product trial, suggesting its primary role is not to convey information about existence, attributes or match values. In the persuasive stream, advertising is usually treated as a taste shifter in utility, and there is usually no specific theoretical justification for its inclusion in the utility function.

Group Inc. and Dish Network Corp., the two biggest satellite TV providers in the US, now offer direct access to chosen households to whom a 30-second ad spot can be targeted (see O'Connor 2014). These ad spots can be bought in real-time by advertisers via "programmatic" ad exchanges (essentially, computer-mediated markets where TV network inventory is sold via auction), thus facilitating a high degree of dynamic, behavior-based targeting (Peterson and Kantrowitz 2014).

The co-determination of advertising and products is key to the targeting problem and for valuing inventory in such markets. Assessing the impact of targeting is predicated on measuring complementarities encapsulated in the cross-partials of utility because that reveals the subset of consumers for whom incremental advertising increases the marginal utility from product consumption. Viewing targeting through the lens of a model with complementarities also changes typical intuition about ad-targeting. As an example, the conventional wisdom is that those who do not like advertising ($\frac{\partial U}{\partial A} < 0$) should not be targeted. However, it is possible that some consumers get negative utility from advertising, but that advertising still increases their marginal utility from consumption (i.e., $\frac{\partial U}{\partial A} < 0$, $\frac{\partial^2 U}{\partial A \partial Q} > 0$), and so should be targeted. Becker-Murphy give the example of fashion- and fitness-related ads that make some consumers feel worse about themselves due to unflattering peer comparisons, but still cause them to buy more cosmetics or fitness-related equipment.

The key to such an analysis is disaggregate micro-data that tracks both ad consumption and purchases. Until now, such data had not been easily available. We leverage access to a new dataset that tracks both the exposure to and consumption of TV advertising by a large panel of households, along with all the purchases made by those households of products in the advertised category. The data are collected by a large market-research company. The purchase data records the products bought, day of purchase, their price, package size, number of units purchased, brand, and manufacturer information for each household. The TV advertising data are recorded down to the minute of the exposure for each household. In addition, the brand associated with each advertising exposure is recorded, enabling us to track the sequence of purchases and ad exposures for over a year for a given household. There are over 100,000 purchase occasions and about 1.5M advertising exposures captured in total. Importantly, the data also record a variable that tracks the fraction of an ad that was played on the TV screen, conditional on an exposure to a TV ad. This variable equals one if the entire commercial was displayed on screen, and is a fraction if the consumer changed the channel or turned off the TV during the commercial. Henceforth, when we refer to "advertising consumption," we refer to this viewing variable. This viewing variable is novel because it reflects more clearly consumer demand for ads and suffers less from the endogeneity issues that are often problematic when looking at ad exposures in a non-randomized setting. Such endogeneity arises when firms set prices and advertising expenditures simultaneously or target advertising to consumers who tend to buy a lot: in the television case, by targeting specific time slots or commercial breaks of shows watched by specific demographics, for example. In our data, conditional on being exposed, the consumer chooses how much of the ad to consume. Thus, we are able to focus our analysis on a component of advertising consumption over which consumers have agency. This is a distinguishing feature of these data compared to traditional "single-source" advertising panels.

We first use the data to test for evidence that advertising is complementary to consumption. An implication of the model proposed by Becker and Murphy is that since advertising enters the consumer's utility function along with other goods, advertising must satisfy the symmetry conditions of utility theory. In particular, this implies that complementarities must go both ways. More ad consumption should raise the demand for product consumption *and* greater consumption of advertised goods should raise the marginal utility from advertising. This is a testable implication of the model.

The main concern in implementing the test is that unobservable tastes that cause individuals to buy more of a product may also cause them to view more ads for the product. We leverage the richness of the panel data to control flexibly for such unobserved heterogeneity. In a battery of specifications, we find that higher purchase quantities increase ad consumption on the margin. To address concerns related to time-varying omitted variables, we leverage the variation observed in the data on the product prices faced by a given household over time. Under the exclusion restriction that these prices do not directly affect the utility from ad-skipping, the observed covariance in the data between low prices paid in the past and future ad-skipping rates identifies complementarities (we discuss our identification strategy in more detail later in the paper). To interpret these effects, we then develop a parametric, discrete-continuous model of demand along the lines of Wales and Woodland (1983); Kim, Allenby and Rossi (2002); Bhat (2005); and Lee, Kim and Allenby (2013), which we estimate jointly with a model of ad-choice. The econometric model allows for complementarities, but does not impose them. The estimates from the model suggest complementarities between advertising and consumption and show significant heterogeneity in these effects across households.

To assess the effects of advertising, we simulate the response to a change in ad exposures, tracking changes in advertising and product consumption in response. We find different implied take-ups of ads across various types of consumers, and find larger effects on purchase incidence and quantity purchased amongst those with higher take-ups, underscoring the need for a precise way of handling endogenous compliance with advertising. Motivated by the "addressable" future of TV ad-markets in which targeting advertising on the basis of ad-viewing and product purchase behavior is possible, we then use the model and estimates to simulate a series of counterfactuals. We simulate how demand, welfare, and profits would change if an advertiser could target ads to consumers (a) on the basis of anticipated skipping behavior (which in the presence of complementarities indirectly selects high demand-consumers); (b) on the basis of the full model of ad-and-product demand; and (c) on the basis of the full model of ad-and-product demand while also implementing targeted first-degree price discrimination. We find that profits are higher under all ad and price targeting scenarios considered, but that targeting on the basis of ad-viewing alone makes up about 16% of the total potential increase in profits, suggesting the value of this policy for advertisers. We also find that net consumer welfare can rise in the new targeted environments, primarily due to the increased surplus accruing to high-volume consumers. These results suggest that it may be possible that firms and consumers are both better off in the new addressable TV environments, though there is considerable heterogeneity across brands on this dimension.

The literature on advertising is voluminous; please see Bagwell (2007) for a comprehensive review. Within this stream, this paper is most closely related to a smaller sub-literature that has empirically assessed the mechanisms by which advertising works. A number of studies in this area have provided evidence for an informative role for advertising by testing whether new or infrequent users respond more to ad exposures compared to established or frequent users (e.g., Ackerberg 2001 on consumer response to ads for yogurt brands; Simester et al. 2009 on consumer response to catalogs mailings; and Tellis et al. 2000 on consumer response to health care referral services). Indirect evidence for persuasive advertising is reflected in the fact that consumers seem to respond to ads that feature strongly persuasive content with few informative attributes (e.g. Bertrand et al. 2010 for experimental evidence on non-informative features of direct mail). As for the complementary view, to the best of our knowledge this paper is one of the first to provide evidence for this mechanism by testing for an implied positive effect of product consumption on ad consumption. In a review of the recent empirical literature, DellaVigna and Gentzkow (2010) note that this "prediction has received some support from laboratory experiments by psychologists (Ehrlich et al. 1957, Mills 1965), but [they] are not aware of any empirical tests from the field."

The paper is also related to an empirical literature on ad-avoidance (e.g., Siddarth and Chattopadhyay 1998, Wilbur 2008, Bronnenberg et al. 2010, Schweidel and Kent 2010, Teixeira et al. 2010, Deng and Mela 2015), though these papers have not explicitly tested the implications of a model with complementarities using micro-data on such avoidance. Outside of advertising, the emphasis here on endogenizing ad consumption jointly with product consumption has some parallels with a literature in labor that has emphasized the endogenous nature of time allocations in models of labor supply (e.g., Biddle and Hamermesh 1990). Similar to the emphasis here on building a formal model of ad consumption, this literature advocates building formal structural models of the time-allocation decisions jointly with labor supply decisions to interpret effects and to evaluate the effect of counterfactual policies. Other papers that are directly relevant to the econometrics and other details are discussed within the body of the paper where relevant.

The rest of the paper is organized as follows. Section 2 discusses issues related to the measurement of advertising effects in the presence of ad-choice in more detail. Section 3 introduces the dataset used for the empirical application and presents evidence of complementarities. Section 4 formalizes a model that allows for complementarities. Sections 5 and 6 present the estimation and simulation results. Finally, Section 7 concludes.

2 Advertising As a Choice: Econometric Implications of Endogenous Non-Compliance

Our approach to measuring advertising effects is to (i) develop and estimate a structural simultaneous equations model of the decision to consume products and advertising and (ii) assess advertising effects through the lens of this model. This section explains in more detail why a model of this sort is useful to assess causal effects of advertising in settings in which advertising is a choice variable for the consumer. We first discuss why randomization alone may not be sufficient to measure advertising effects with policy-relevant economic content in such settings, and then discuss the implications of consumer ad-choice for advertiser and TV networks' policies.

The main econometric challenge in measuring causal effects of advertising in settings with skipping is non-compliance. Using the terminology of program evaluation, if one views advertising as the "treatment," randomization alone cannot measure the *treatment effect* of advertising for all sub-populations of consumers of interest. To set up some notation, denote an individual by i and let d_i be an indicator of whether i consumes an ad associated with a brand. Let y_{i0} be i's outcome if no ad is consumed, and y_{i1} the outcome when the ad is consumed. Define the treatment effect of advertising, θ_i as,

$$\theta_i = y_{i1} - y_{i0} \tag{1}$$

If d_i is randomized to consumers, we can measure the average effect of the ad across all i,

$$ATE = \mathbb{E}\left[y_{i1} - y_{i0}\right] = \underbrace{\mathbb{E}\left[y_{i1}\right] - \mathbb{E}\left[y_{i0}\right]}_{\text{linearity}} = \underbrace{\mathbb{E}\left[y_{i1}|d_i = 1\right] - \mathbb{E}\left[y_{i0}|d_i = 0\right]}_{\text{randomization}} = \mathbb{E}\left[y_i|d_i = 1\right] - \mathbb{E}\left[y_i|d_i = 0\right] \tag{2}$$

where the first equality obtains because of the linearity of expectations and the second from the fact that the treatment d_i is randomized, and therefore $d_i \perp \{y_{i0}, y_{i1}\}$. Now suppose that consumers actively choose to see the ad conditional on being randomized into the ad condition (the situation considered here). Let $\tilde{d}_i$ denote whether i was assigned to the ad condition, and d_i denote whether i actually *chose* to see the ad. Those with higher θ_i (for example, those that value the brand more) will tend to choose $d_i = 1$ (i.e., will be more likely to choose to consume the ad). From (1), these high θ_i individuals will tend to have higher potential outcome differences. Because of this differential compliance, even though $\tilde{d}_i \perp \{y_{i0}, y_{i1}\}$, now d_i is no longer independent of $\{y_{i0}, y_{i1}\}$. Essentially, the "treatment" is no longer randomized and the decomposition in (2) no longer obtains. Since the set of individuals who see the ad are different from the control, even with randomization we cannot measure the treatment effect of ad consumption.[2]

Non-compliance was first recognized in the medical field as a statistical problem that confounded estimation of treatment effects because some patients assigned the treatment refused to consume the drug, or dropped out of the treatment intervention. The concern is that those who drop out are different from those who stay because they perceive the benefits of treatment to be lower. Medical researchers solved this problem by "double blinding," so the set of treated patients who refuse to comply with the treatment are not aware if they are assigned the treatment drug or the placebo. When non-complying patients do not know they are in the treated or control groups, there is no reason to believe that non-compliers are more averse to treatment than compliers, so this does not confound the measurement of treatment effects. However, in advertising situations, the double blinding strategy is not feasible because a consumer always sees an ad before deciding to skip it or to see it fully. Thus, ad consumption *per se* cannot be randomized.

Why would an advertiser or a TV network care about non-compliance? Randomization does identify the intent to treat effect of advertising (ITT) (the average effect of being assigned to the ad condition) even with non-compliance, and the ITT is a sufficient metric with economic content for some questions. In particular, the ITT is sufficient for assessing the return from an advertising campaign run on the entire population. If an advertiser randomizes viewers into an ad campaign and wishes to assess overall campaign effectiveness, it is sufficient to know the net profit from those assigned to the campaign relative to those not, without having to know the differential intervention effects for individuals with different compliance types.

In other situations − most notably, those involving the targeting of advertising to individual consumers − the advertiser may care about knowing differential ad-responses for individuals as well as their anticipated compliance. As we explained in the introduction, such targeting at the individual-level is increasingly becoming common in TV ad markets. For individual-level targeting, a model of whether the targeted viewer will comply, as well as an assessment of the actual treatment effect of the ad is important. For instance, an advertiser may decide to target a given digital video ad unit to consumers who are more likely to watch it, or to those for whom the response from the ad is highest. For this, the ITT alone is not sufficient. In other situations, the advertising firm may care not just about the total number of units sold in response to advertising, but about the *composition* of buyers *per se*. Credit-cards, insurance and other financial product markets are leading examples, because the cost curve faced by the firm is a function of the composition of customer types, and not just the total number of card or policy holders. Hence, credit-card companies and auto insurance firms − two sets of high-spending TV advertisers in

[2] To be clear, the treatment effect of ad *exposure* is still identified by randomization under this situation, but the treatment effect of ad *consumption* is not.

the US − care about the *type* of customers who respond to their advertising because they would like to avoid attracting high-cost, high-risk agents to their customer pool. This requires knowing who out of the targeted sub-population will respond to the advertising.[3] In these contexts, we would like to learn the full distribution of advertising effects, not just the mean effect of ad exposure as in Equation (2). In the endogenous compliance case with heterogeneous consumer response, it is difficult to characterize which sub-population will consume the ad, the distribution of treatment effects for all sub-populations of interest, or even the mean treatment effect for all consuming sub-populations, from randomization alone.[4] All of these policy-relevant questions are difficult to address without a well-posed model with heterogeneity that characterizes these sub-populations and articulates the effects.

From the TV network's perspective, ads that are less likely to be skipped may be favored all else equal, because they could reduce the chance that consumers may switch away from TV during commercials. Similarly, a TV ad network may be willing to entertain price discounts on ads targeted to consumers who are less likely to skip them. More generally, a TV network that would like to assess the price to charge advertisers for specific sub-populations of its viewers would find it useful to know which subsets of viewers and of what type actually see the ads targeted to them, and what the effect on the advertiser's sales and revenue were from each subset's exposure to those ads. This requires measurement of actual treatment effects. In other situations, it may be of separate interest to a firm to measure what proportion of consumers of a given type who are assigned to an ad actually view it, so as to measure consumers' taste for privacy, or to assess their nuisance value of advertising. Or it may be of interest to a researcher to measure the efficacy of advertising *per se* ("what would happen if an agent saw the ad") as opposed to assessing the effectiveness of an ad-campaign ("will the campaign work when some viewers could plausibly skip ads"?). For situations such as these, the ITT metric and the randomization strategy alone are insufficient. Finally, a related question is why do some sub-populations respond and others not? This requires recognizing that the decision to take up treatment is a function of anticipated gains from the treatment, along with a clearly understood mechanism for why heterogeneous consumers decide to consume advertising.

Taken together, in our view, a well-posed and empirically realistic model of ad consumption is important to interpret advertising effects.

[3]Evidence in the literature suggests a link between those who respond to advertising and risk in such markets. Using randomized trials on direct-mail advertising, Ausubel (1999) documents that customer pools resulting from credit card offers with inferior terms (e.g., a higher introductory interest rate, a shorter duration for the introductory offer) have worse observable credit-risk characteristics and are more likely to default than those drawn in by solicitations offering superior terms.

[4]To see this, note that when advertising consumption is an explicit choice, the treatment assignment of individual i, $\tilde{d}_i$, should properly be viewed as an instrument for d_i, and randomization facilitates an instrumental variables (IV) estimator of the effect of advertising. Following Imbens and Angrist (1994), with heterogeneous treatment effects, IV measures a local average treatment effect for a specific sub-population of compliers − i.e., a set of consumers that are induced by assignment to the ad-condition to change their decision to consume ads. Unfortunately, this sub-population cannot be characterized without additional assumptions, nor can the measured effect be extrapolated to any other sub-populations of interest.

By placing advertising consumption in the same footing as product consumption, Becker and Murphy's model provides a framework to handle ad and product choice with a clear link to micro-foundations. This theoretical setup has to be modified in four ways when confronting household-level panel data on TV ad consumption (like in our empirical application). First, it needs to be augmented to allow for stock effects of advertising (as opposed to purely flow effects) in order to handle the carryover effects of advertising that have been documented in past empirical work (e.g., Naik et al. 1998). Second, we need to have a definition of "advertising consumption" that can be sensibly interpreted as reflecting consumer demand for TV ads. Third, TV ads are at the brand-level, and hence the model has to be modified to allow for choice over brand-level quantity and advertising consumption. Fourth, purchase and ad-viewing decisions may be sequentially rather than simultaneously determined in TV decision contexts where the product purchase and ad consumption decisions are separated in time and location.

3 Data Description

The dataset used in our empirical analysis comprises a panel of household-level matched purchase and advertising data from a large sample of households in a Western European country. As part of the conditions for acquiring the data from our corporate partner, the identities of the product category and origin country cannot be revealed. Non-confidential details of the context and category that are relevant to interpreting the empirics are described below.

Data Collection The TV viewing data collected by the market research firm forms the standard for the TV ratings for the broadcasting industry in the country. The country in question has been a pioneer in using and developing technology in the field of TV audience measurement. The firm maintains a panel of households whose TV consumption behavior is tracked. Households receive a financial incentive for participating in the panel. Upon recruitment, households are provided guidance on how to use the tracking device and given a user manual. Households are chosen so the sample is representative of the population of the country. TV consumption of the household members is measured using a modern version of a "people-meter," which consists of a measuring device connected to the TV, a display and its own remote control. As a user starts watching TV, s/he declares his identity by pressing an individual presence button on the remote. The device then monitors and records to the precise second every time the TV set is turned on or off or the channel is changed, as well as any other types of use, such as games consoles and electronic program guides. Time-shifted viewing on digital recording devices, such as hard drive receivers or DVD recorders is also collected. The specific program or ad that is played on screen is identified by audio-matching. An identifier (essentially an audio watermark) is inserted into each content that is broadcasted. The device picks up and stores the identifier for each user. Each night, the data from each household are uploaded to the company's server. The identifier is matched with the broadcasting schedules in order to identify the content played. TV audience measurement in most developed countries is now done this way.[5]

An important feature of the data is that in addition to tracking the panelists' TV viewing behavior, the data also records the households' in-store purchases for a set of product categories. Households agree to scan their purchase receipts periodically, and those data are transmitted back to the firm. This enables creating a profile of ad and product consumption over time. Purchases are recorded at the household-brand-day level and ads are captured at the household-brand-exposure level.[6] Demographic information about the households, including the number of family members, number of children, and average income and education levels of the head of household is also available. Finally, the data also records a variable that tracks the fraction of an ad that was played on the TV screen, conditional on an exposure to a TV ad. This variable equals one if the entire commercial was displayed on screen, and is a fraction if the consumer changed the channel or turned off the TV during the commercial. We treat this variable as representing "ad consumption." We have access to advertising and product purchase data for one category. Data on views of non-ad TV is not available. Various audits are conducted periodically by the firm to verify its integrity.

Product Market The product category is a fast moving consumer packaged good (CPG) with established brands that is primarily sold in brick and mortar stores. More than 80% of the product is sold via big-box retail, while the remaining 20% is sold in smaller stores like convenience stores and gas stations. Consumption of the product in the country is one of the highest in the world on a per capita basis (higher than the US). Retail pricing in the category is characterized by a steady base price, punctuated by frequent temporary price promotions ("sales") and seasonal discounts, inducing a high level of local price variation, as is typical of CPG-retail.

Advertising The category is heavily advertised on TV. Most of the TV advertising in this category is not directly informative (providing details of product attributes, sales or prices). Instead, most ads are focused on associating the brand with the pleasure of product consumption, documenting scenarios where individuals consume the product in a variety of settings, and containing humor, stories and other narratives ("image advertising"). A unique feature of the market is that all television ad-buys by firms are at the *national*-level, with no tailoring or targeting at the regional level.[7] All regional TV is publicly

[5] For an example of such data, see `http://www.finnpanel.fi/en/tv.php` (not necessarily the same as the country in our sample.) Examples of the types of devices used here include the UNITAM meter by Nielsen Media Research (Unitam, 2009) and the RapidMeter by Kantar (Kantar, 2012). For a video demonstrating how Nielsen's system works, see `https://www.youtube.com/watch?v=jYrVijea0UM`

[6] Per the recommendation of the company sponsor that provided the data, we define a brand using the variable denoted "umbrella brand" in the dataset. This is also the level at which the advertising data is collected.

[7] TV ads also cannot be targeted to an individual − all individuals watching the same show see the same ad.

funded and does not contain ads. This feature of the market implies that there exists significant local price variation that is not coordinated with the content or the incidence of advertising on TV. Later, we will exploit this feature for identification.

Appropriateness of Data and Context Brand-oriented advertising on television represents a good social laboratory to assess Becker and Murphy's theory, as the complementarity mechanism they proposed was originally motivated to explain the prevalence of non-information related, branding-and-lifestyle oriented ads in the marketplace. CPG firms typically serve such "image advertising" mostly on TV. Other digital formats like search and display advertising are less suited because most advertising in those media are "performance-oriented," featuring coupons, price-reductions, or products recently viewed, rather than image-oriented content. Unlike such digital media, it is more difficult to collect data that tracks both ad exposure on TV and matched offline purchases at individual-level granularity. This data has the advantage of tracking both. Two disadvantages of the data are, (1) the purchase data can be matched only at the household and not the individual-level, and (2) the technology does not account for whether the panelist was paying attention to the advertisement or if the television was muted when the content was delivered. The latter introduces measurement error into the ad consumption variable. We are not aware of a measurement technology that solves this problem for ad consumption. In the absence of being able to track a consumer's brain activation explicitly, the researcher will always invariably have to contend with confronting some error in the measurement of "attention."[8] Our approach to handling the measurement error is discussed later in the paper.

3.1 Descriptives

Table 1 presents aggregate summary statistics for the purchase and advertising data. The data runs from June 14, 2010 through December 31, 2011 covering about 6,500 households (i.e., a balanced panel with $T = 557$ days). 58 distinct brands are purchased. We focus our analysis on the 11 brands with the largest market share. There are over 100,000 purchase occasions and about 1.4M advertising exposures captured for these brands.

Table 2 provides summary statistics on the distribution of purchases for those households who made at least one purchase in the category. The median household made 13 purchases and bought 4 different brands in the category (mean inter-purchase time of 29 days). Figure 1a shows the distribution across households of the total number of purchases, and Figure 1b shows the distribution across households of the total number of brands purchased, over the course of the panel. There is significant heterogeneity in both brand preferences as well as the total frequency of purchase.

[8] "Eye-tracking" data (e.g., Teixeira et al. 2010) holds promise in improving measurement of advertising attention, but we have not seen these data collected at scale in field-settings and matched to purchases.

Table 1: Summary Statistics

Full Panel Starts	6/14/10
Full Panel Ends	12/31/11
Brands	11
Households	6,437
Purchase Occasions	117,516
TV Ad Exposures	1,445,389

Table 2: Across Household Variation in Purchases

	N	Min	Median	Mean	Max
HH Purchase Occasions	6,272	1	13	19	304
HH Brand Count	6,272	1	4	4	11

Note: Reported for the 6,272 HHs who made at least one purchase.

Figure 1: Distribution of Purchases and Number of Brands Purchased by Household

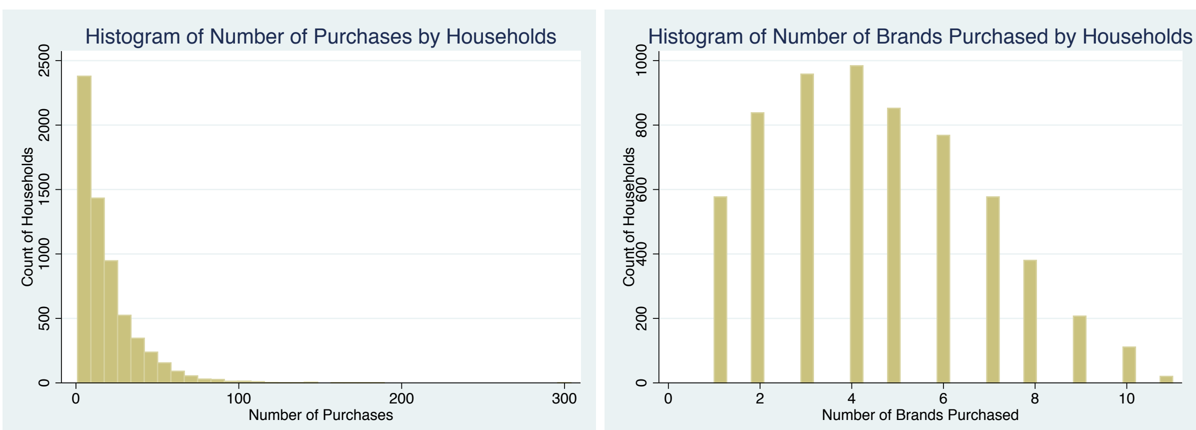


Figure 2: Distribution of Ad Exposures by Household

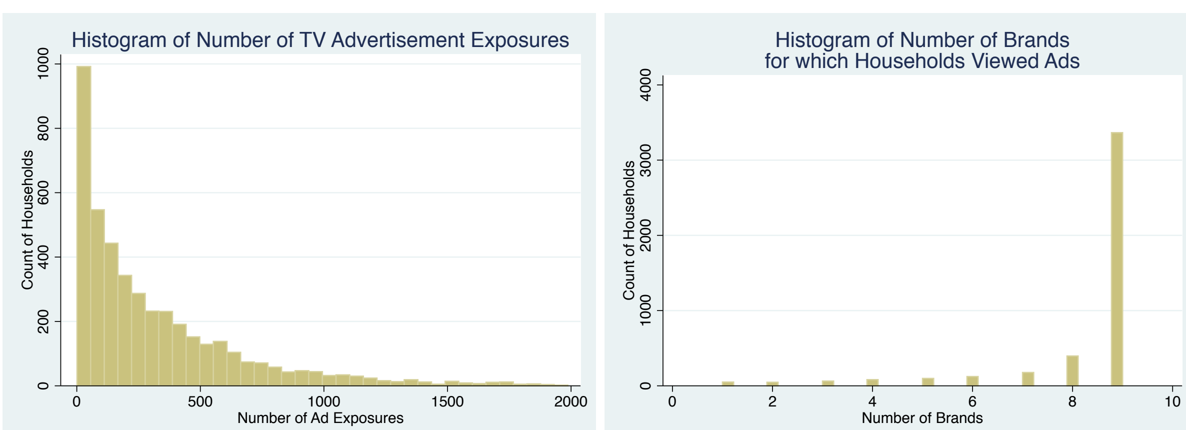


Table 3: Across Household Variation in TV Advertisement Exposures

	N	Min	Median	Mean	Max
HH TV Ad Exposures	4,401	1	199	328	3,808
HH Brand Ad Count	4,401	1	9	8	9

Note: Reported for HHs who viewed at least one TV advertisement.

Table 4: Variation in Daily Purchase Quantity, Conditional on Purchase

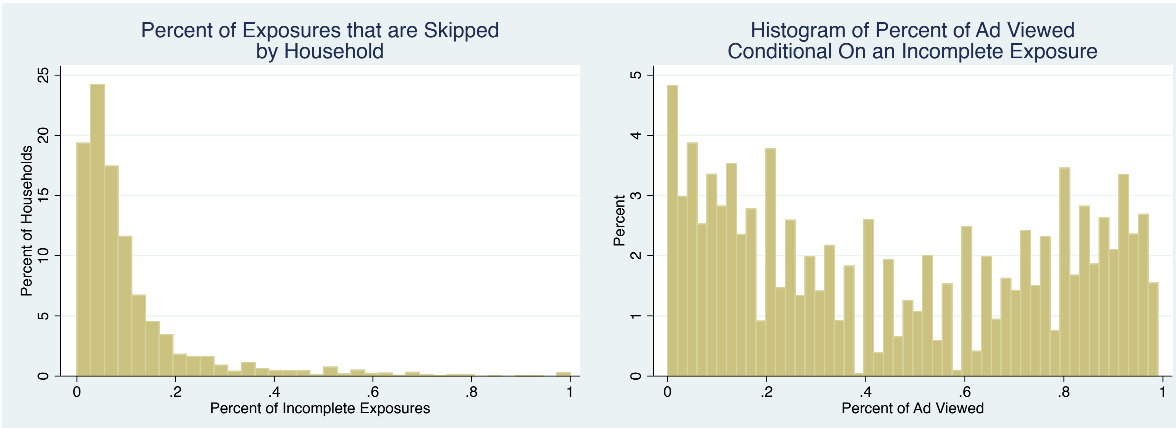
	N	Min	Median	Mean	Max
Purchase Quantity	117,516	119	1,617	2,015	42,000

Table 3 provides the analogous summary statistics for advertisement exposures. The median household views 199 TV ads and views a TV ad for 9 different brands in the category (mean = 0.6 exposures per day). Figure 2a shows the distribution of advertisement exposures across households. There is a spike in the distribution at the lower end, but there is extensive variation in the number of exposures across households. Figure 2b summarizes the number of brands for which households viewed at least one advertisement. Two of the 11 brands in our analyses do not advertise, so the maximum number of advertised brands is 9.

Turning to advertising consumption, Figure 3a shows a histogram of ad-skipping rates across households. We define a household's ad-skip rate as the proportion of that household's total ad exposures over the observed length of the panel that is not watched to completion (i.e., the proportion of exposures for which the corresponding ad consumption variable is less than 1). The histogram shows large heterogeneity in skip-rates across households with some skipping more than 60% of the ads to which they are exposed. The median household skips about 10% of the ads it sees. Ignoring the household-level variation, if we look at all ad exposures across all households, we find that about 5% of the ads are skipped. Figure 3b shows a histogram of the variation in ad consumption for the subset of skipped exposures (there are about 72,000 such observations, i.e., $\approx$ 5% of 1.4M). We see wide variation in how much of an ad is viewed conditional on the decision to skip it.

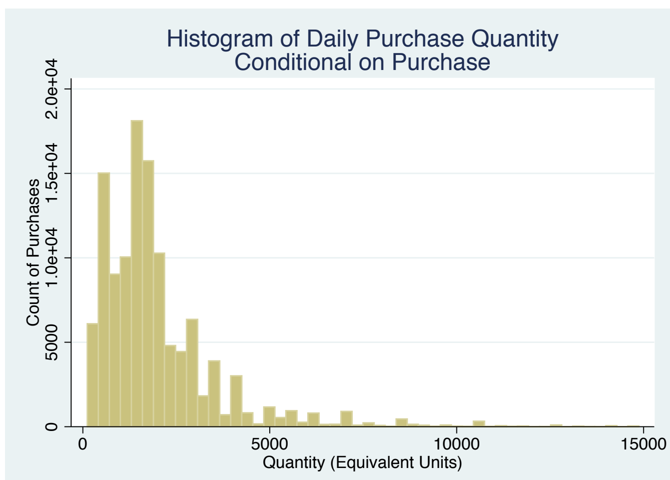
Table 4 now reports summary statistics for the quantity purchased on a given purchase occasion; Figure 4 shows its distribution across households. We define the quantity purchased as the number of units bought in a day of a given brand times the equivalent volume of that unit (i.e., the total package volume of all purchases of brand j made by household i on day t expressed in equivalent units). Because of the company sponsor's desire to remain anonymous, the amount of a product purchased is reported in units of equivalent volume, without specifying exactly what scale these map to (we cannot convert it to say grams, pounds or liters). The mean equivalent volume purchase of a brand is about 2,000 units, and there is extensive variation in purchased quantity across purchase occasions. Our test of the theory boils down to a test of whether the observed variation in quantity purchased can explain the observed

Figure 3: Distribution of Ad Skip Rates and Percentage Watched



variation in ad-skipping.

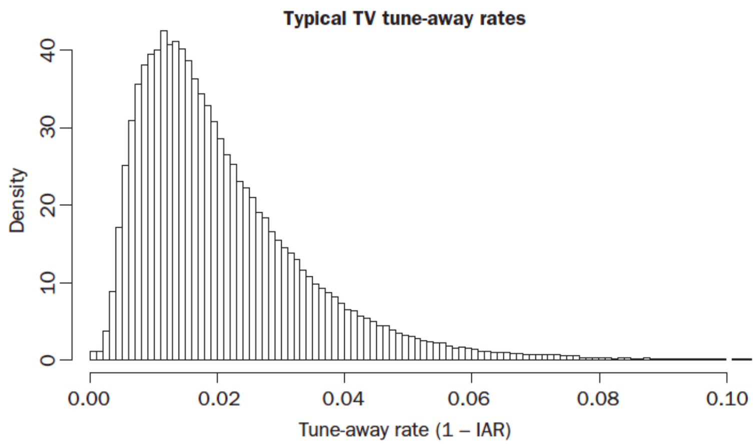
Figure 4: Histogram of Daily Purchase Quantity, Conditional on Purchase



Before concluding this section, we briefly discuss two aspects of our ad consumption data. First, the 5% ad-skip rate reported above warrants some discussion as it may seem small compared to casual intuition about ad-skipping, especially in relation to online ads.[9] The skip-rates we observe in our TV data are consistent with those reported in the few academic papers we know of that have access to data on household-level TV ad-skipping rates. For example, Bronnenberg et al. (2010) analyzes TiVo log-data and finds that TiVo households fast-forward through 6.5% of the 46,620 ad exposures in their

[9]Typically, reported skip-rates of TV ads are lower than skip-rates of online ads. This difference may arise because the effort required to skip an ad online (ignoring a banner ad or clicking to skip a YouTube TrueView ad) is generally less than the effort required to skip a TV commercial (changing the channel and monitoring when to return to the program). Some advertising executives we spoke to stated that it could be because the passive default option for an online consumer is to ignore the ad, while the action that involves some effort on his part is to click on it. In television advertising, this is reversed: the passive default option for the consumer is to view the ad, while the action that involves some effort on his part is to change the channel.

Figure 5: Distribution of Ad Tune-away rates in Dish Network Set-Top Data (reproduced from Interian et al. (2009a), and Zigmond et al. (2009))



data. Similar rates are reported in research from Google using TV set-top data. Figure 5 reproduces ad-skipping rates reported in Interian et al. (2009a), and Zigmond et al. (2009) based on data acquired by Google from the DISH Network in the US. This data describes the second-by-second tuning behavior of television set-top boxes in millions of US households. Analyzing 182,801 ad placements, they report mean "tune-away" rates – defined as the proportion of the audience that starts viewing an ad and tunes away from it without watching it completely – of 1%-3%. Our data are consistent with these numbers.

Second, we may worry whether ad consumption can be meaningfully studied in the absence of TV viewing data. In more recent research using data from TiVo logs, Deng and Mela (2015, Table 1) analyzes the extent to which contextual factors explain the variation in ad-skipping across exposures.[10] They find that each of these factors explain less than 1% of the observed variation. Rather, the bulk of the variation in ad-skipping is explained by household fixed-effects (20.4%) and past observed propensity to skip ads (11.9%). In addition, if household fixed effects are replaced by a set of demographic variables, the demographics account for only 3.2% of the variation in ad-skipping, suggesting that unobserved household-level heterogeneity is significant in explaining ad-skipping. In our data, we do not observe TV network or show information, but we do observe data on household characteristics. Like Deng and Mela, we explore what percent of the variation in ad exposure and ad skipping can be explained by these characteristics by regressing household exposures and skip rates on the set of observed consumer characteristics (Appendix A, Table 15). The observed characteristics explain little of the variation in ad exposures and skip rates across households.

[10]Variables considered include the brand of the ad, show genre, network in which the ad airs, product category, location of the commercial break within the show and the slot within the break, day of week and hour of show.

These stylized facts from previous research provide some face validity for our analysis, which focuses on household preferences over product and advertising consumption as the main explanation for ad-skipping variability, as opposed to show, network and other TV-environment specific characteristics. In our set-up, the heterogeneity in household skip-rates will be explained by preferences over ad consumption and by the observed quantities consumed of the advertised brands. The co-dependence of ad-skipping rates on the quantity consumed of the advertised products is a novel feature of our empirical model that provides a mechanism for the observed state dependence reported in product and advertising consumption in past studies.[11]

3.2 Testing the Model

The implication of a joint model of complementarities is that more consumption of an ad induces more consumption of the product, *and* more consumption of the product induces more consumption of the ad. That advertising causes exposed users to consume more of the product is also predicted by other mechanisms for why advertising works and is not novel to the complementarities theory. However, the opposite prediction − that product consumption causes users to consume more of the product's ad on the margin − is a unique implication of the complementarities mechanism. Hence, we base empirical support for the theory on whether we detect a positive effect of ad consumption on product consumption in the data.[12]

The main challenge in implementing the test is the confounding effect of common variables that shift both ad and product consumption. For example, it could be that larger households are more likely to view more advertisements because they are more likely to own multiple TVs, or to watch family oriented shows that feature more of the product's advertisements. At the same time, larger households may be more likely to buy more because they have a higher demand for the product. Household composition is just one of many observable household characteristics that could potentially be related to both advertising and purchase behavior. While we can control for these observable characteristics, there may also be unobservable characteristics of households that may generate spurious correlation. In general, any correlated unobservable that affects both the propensity to buy a brand and view ads of that brand can confound the test. A separate concern is the extent to which observed skip rates are

[11] For instance, Interian et al. (2009b) reports an interesting correlation that the more often viewers have seen an ad over the last month, the less likely they are to tune away. This correlation can be explained if ads have a positive effect on quantities purchased, and quantities in turn have a positive effect on ad views.

[12] What could be the psychological underpinnings of such complementarity? While we cannot provide data-based support for more micro-explanations, we conjecture one reason consumers may watch more ads of products purchased recently may derive from "licensing," wherein users watch ads of others consuming the product as a way to justify to themselves their own consumption (Shafir, Simonson and Tversky, 1993). Behavioral researchers such as Prelec and Lowenstein (1998) have pointed out that such behaviors are likely when consumption of the product evokes a sense of guilt (e.g., expensive luxury items, junk food, indulgent goods). Another explanation is that watching ads of products purchased recently may provide utility to users from re-living the felt-utility from enjoyable past consumption of the product. Past literature (e.g., Lowenstein and Elster 1992) has pointed out that reliving and contemplating past experiences is a source of significant utility for human beings.

understated by measurement error. As noted before, it could be that households lose attention or look away when an ad is playing, a form of non-consumption that is not captured by the TV ad measurement technology. This produces measurement error in the y-variable (ad consumption), which can reduce the precision of the estimates or induce a bias if it is systematically correlated with the included quantity variable. Finally, we may also be worried about attenuation bias in the x-variable (purchase quantities). Even though quantities are captured without error in the data, measurement error may arise because we are using purchase quantity as a proxy for product consumption, the "correct" x-variable implied by the theory.

We address these issues in the following way. First, we do not use the variation in consumer *exposure* to advertising to inform our test. This would be problematic because exposure to an ad is determined by consumers' show-preferences and by firms' decisions to target their ads to shows. The worry in using this variation is that consumers who tend to buy more of the product may also tend to watch more of the shows that ads are served on, so any positive dependence between quantity and ad exposure could reflect these unobserved tastes and firm's advertising supply decisions. Instead, our strategy relies on the fact that, conditional on being exposed to an ad, consumers have agency over how much of the ad to watch. Therefore, by using the percentage of the ad watched conditional on exposure as our dependent variable of interest, we assess complementarities on a variable that more closely reflects consumer demand for advertising.

To address the confound due to correlated unobservables, we leverage the panel aspect of the data to use only the within-household, over-time variation to test for the effect of purchase quantities on ad consumption. The use of the within-variation controls for any correlated time-invariant unobserved heterogeneity. We also include week fixed effects to control for seasonality (e.g., consumers buying more of the product and watching more of the ad on the margin during holidays). This forms our "base" set of results.

There may remain concerns about individual-and-time period specific omitted variables that cause some individuals to buy more of the product and watch more ads. Addressing this requires access to instrumental variables. For this, we leverage the fact that households in our data are exposed to significant variation in prices for the products over time (we document this below). We impose an exclusion restriction that product prices affect quantity purchased, but do not directly affect the percentage of ads watched. We believe this exclusion restriction is reasonable and discuss how the institutional setting facilitates the identification. In essence, we ask: "All else equal, are more of a product's ads consumed if the household faced low prices in the past?" Under the maintained exclusion restriction, this covariation identifies the effect of past quantity purchased on subsequent ad consumption. This strategy is analogous to Gentzkow's (2007) approach for measuring substitution and complementarity between online and offline newspapers, and is discussed more formally in Fox and Lazatti (2014). The instruments will also correct for attenuation bias. However, because *a priori* we do not have a view on the direction of the omitted variable bias in our fixed effects estimates, we view our instrumental variable regressions as a sensitivity check on those estimates. We use the IV estimates to sign the extent of the bias. We argue on the basis of our results that the fixed effects estimators are downward biased, and hence represent a conservative test for the complementarity.

3.2.1 Fixed Effects Regressions

The regressions below treat the current advertising consumption as the dependent variable and cumulative product consumption of the advertised brand as a regressor. The dependent variable a_{ijrt} is the percent watched (range: 0-1) of the r^{th} ad exposure for brand j watched by household i on day t, and the right hand side variable, Q_{ijt}, is the cumulative quantity purchased by i of brand j over the last two weeks.[13] The θ_0-s are household, brand, and week fixed effects. Each row in the regression is a household-brand-day-exposure combination,

$$a_{ijrt} = \theta_{0i} + \theta_{0j} + \theta_{0w(t)} + \theta_1 Q_{ijt} + \epsilon_{ijrt} \quad (3)$$

Table 5 reports the results. Unless stated otherwise, robust standard errors clustered at the household level are reported for all specifications going forward. Looking at column 1, past quantity is seen to have a positive and significant effect on ad-skipping (we report marginal effects below). Column 2 adds day of the week dummies, and finds the results are consistent with the specification in Column 1. For completeness, we also estimate the analogous regression using the percentage of the ad viewed unconditional on exposure as the dependent variable. Here, days in which there are no exposures to a given brand's ad for a household are also included as rows in the data with the value of the dependent variable set equal to 0. Hence, this regression explores the effect of past purchase quantity on both the propensity to be exposed to an ad, and the propensity to view it for longer conditional on exposure. Looking at column 3, we see the effect of past quantity is positive, though this regression is hard to interpret as it mixes the supply and the demand for ads.

[13] We considered cumulative quantity measures ranging from quantity purchased in the preceding day, $\tilde{Q}_{ijt,1}$, to quantity purchased over the preceding four weeks, $\tilde{Q}_{ijt,28}$. We found a positive and significant ($\alpha = 0.05$) relationship between ad consumption and cumulative quantity for $[\tilde{Q}_{ijt,11}, \tilde{Q}_{ijt,15}]$. Outside of this range the relationship between cumulative quantity and ad consumption was consistently positive, but not statistically significant.

Table 5: Regressions of Ad Consumption on Cumulative Quantity

	(1) Percent Ad Watched Conditional on Exposure	(2) Percent Ad Watched Conditional on Exposure	(3) Percent Ad Watched Unconditional on Exposure
$\tilde{Q}_{ijt,14}$	0.0010**	0.0010**	0.0023***
	(0.0004)	(0.0004)	(0.0003)
Observations	1,436,400	1,436,400	29,778,304
HH FE	Y	Y	Y
Brand FE	Y	Y	Y
Week FE	Y	Y	Y
Day of Week Dummies	N	Y	N

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Regression estimated at the household-brand-day-exposure level. The dependent variable Percent Ad Watched records the percentage of the exposure that was watched and ranges between 0 and 1. Columns 1 and 2 are estimated conditional on an exposure. In column 3, the dependent variable is recorded as a 0 for days in which no ads were viewed. $\tilde{Q}_{ijt,14}$ records the cumulative package volume household i purchased of brand j in the 14 days preceding day t, normalized by the average daily purchase volume, $\bar{q}_t = 2{,}015$ equivalent units. Robust standard errors clustered at the household level.

Table 6: Regressions of Daily Ad Consumption on Cumulative Quantity for Households with Different Skip Rates

Column:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent Variable:	% Ad Watched	% Ad Watched	% Ad Watched	% Ad Watched	% Ad Watched	% Ad Watched	% Ad Watched
HH Ad-Skip Rate is >	0	0.01	0.05	0.10	0.15	0.20	0.25
$\tilde{Q}_{ijt,14}$	0.0010**	0.0010**	0.0011	0.0018	0.0009	0.0047	0.0131
	(0.0004)	(0.0004)	(0.0008)	(0.0019)	(0.0047)	(0.0079)	(0.0127)
Observations	1,436,400	1,369,030	558,595	159,242	57,764	23,522	12,327
R-squared	0.041	0.040	0.038	0.053	0.071	0.093	0.111
Mean Dep Var	97%	97%	95%	92%	89%	85%	81%
HH FE	Y	Y	Y	Y	Y	Y	Y
Brand FE	Y	Y	Y	Y	Y	Y	Y
Week FE	Y	Y	Y	Y	Y	Y	Y
Marginal Effect of an Additional Purchase of Various Quantities on Expected Percentage Watched							
Mean	0.10%	0.10%	0.11%	0.18%	0.09%	0.47%	1.31%
25th Percentile	0.04%	0.04%	0.05%	0.08%	0.04%	0.21%	0.59%
50th Percentile	0.08%	0.08%	0.09%	0.15%	0.07%	0.38%	1.05%
75th Percentile	0.11%	0.12%	0.13%	0.22%	0.11%	0.56%	1.54%

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Regression estimated at the household-brand-day-exposure level. The dependent variable Percent Ad Watched records the percentage of the exposure that was watched and ranges between 0 and 1. $\tilde{Q}_{ijt,14}$ records the cumulative package volume household i purchased of brand j in the 14 days preceding day t, normalized by the average daily purchase volume, $\bar{q}_t = 2{,}015$ equivalent units. The regression reported in column 2 only includes observations for households who skipped at least 1% of their ad exposures. Columns 2 through 7 restrict the sample to observations for households with increasingly higher skip rates. Panel 2 reports the expected increase in ad percentage watched from an additional purchase in the preceding 14 days. The marginal effects are reported for different quartiles of the purchase quantity distribution. Robust standard errors clustered at the household level.

To explore the heterogeneity in advertising consumption effects, we repeat the same regression separately for households with different observed ad-skip rates. Table 6 repeats the regression from Column 1 of Table 5 separately for households with mean observed ad-skip rates over the entire data that are greater than 1%, 5%, 10%, 15%, 20% and 25%, respectively. For ease of comparison, Column 1 in Table 6 repeats the results from Column 1 of Table 5. All regressions report the effect of cumulative quantity purchased over the past 2 weeks on the percentage of an ad viewed by a household, conditional on exposure, while including household, brand, and week fixed effects. Although we lose power when focusing on only households that have high skip-rates, the effect of past product consumption is positive for all subgroups of households, and the marginal effect of quantity on ad consumption is higher for those with higher observed skip-rates.

To interpret these numbers, in the bottom panel of Table 6, we also report the effect on ad consumption of an increase in the quantity consumed of the product over the past two weeks for each of these subgroups. To do this, for each subgroup, we calculate the mean, 25th, 50th and 75th percentiles of the quantity purchased by households in that subgroup on days with a purchase. We denote these values as $(\bar{q}, q_{.25}, q_{.5}, q_{.75})$. We report how much ad consumption would change if a household in each subgroup increased its quantity purchased over the last two weeks by these values. For instance, column 7 reports the results for households with an ad-skip rate > 25%. Looking at the bottom panel of column 7, we see that if the quantity purchased over the previous two weeks is increased by $\bar{q}$, households in that subgroup are likely to watch 1.31% more of the brand's ad, conditional on exposure. If the quantity purchased over the previous two weeks is increased by $q_{.75}$, households in that subgroup are likely to watch 1.54% more of an ad, conditional on exposure. Overall, these regressions suggest that past purchase of a brand causes consumers to watch more advertising of that brand on the margin.

3.2.2 Instrumental Variable Regressions

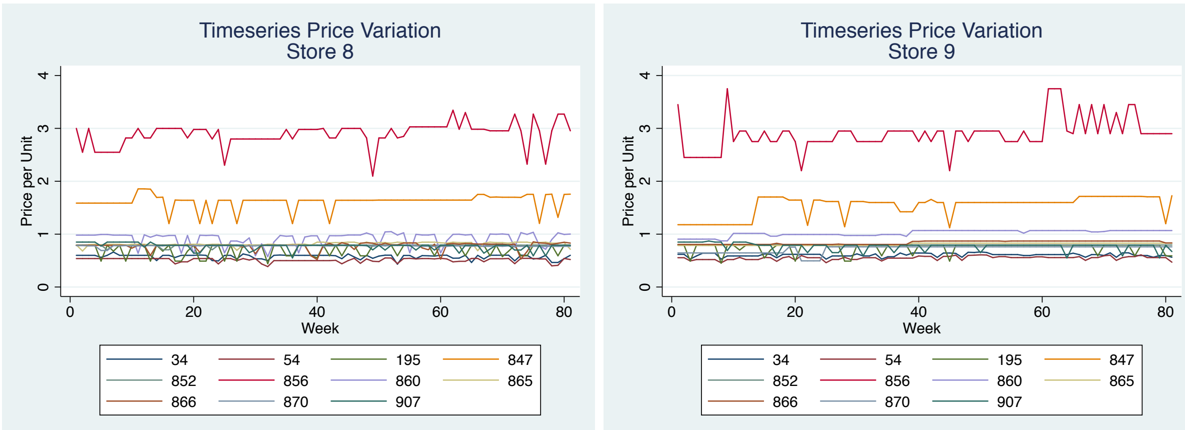
The IV regressions reported below speak to concerns about unobservables that are consumer-and-time-period-and-brand specific that are correlated with recent quantity and to any measurement error in the $x-$variable. *A priori*, one could tell different stories for what the unobservables represent, with differing implications for their correlation with ad consumption and quantities. For instance, suppose a consumer is on vacation and is less likely to skip ads because he has more time on his hands. During vacation, the consumer may also be more likely to buy the product, so then, the estimated effect is upward biased. Another account is that during vacation, the consumer prefers to watch more shows, and is therefore more likely to skip ads. If the consumer is also likely to buy more of the product during vacation, then the estimated effect is downward biased. Separately, measurement error induces attenuation that may be accentuated with fixed effects (e.g., Ashenfelter and Krueger, 1994). This produces a downward bias. But, because we do not know which interpretation about the unobservables is correct, we cannot *a priori* sign the direction of the omitted variables bias. Our strategy will be to infer the nature of the net bias ex-post by comparing the FE estimates to the IV estimates.[14]

Price Variation The IV strategy leverages the price variation in the data under the maintained exclusion restriction. To see what price variation is useful for identification, note that prices vary significantly across brands, stores, and time in the data. However, utilizing the *across-brand* price variation is problematic because some brands may have ads that are watched more, and may also have lower prices, so the correlation between low prices and high ad consumption across brands may simply reflect this association, and not the effect of quantities purchased. Using the *across-store* price variation is also problematic, because the correlation between prices and ad consumption across stores-visited may simply reflect unobserved consumer heterogeneity and not the effect of quantities purchased. For instance, some stores that have high prices may systematically be visited by high income consumers, and these consumers may skip ads more often given their higher marginal costs of time. So, the fact that ad-skipping is lower when comparing high-price store visitors to low-price store visitors, may simply reflect differences in the type of consumers that visit these stores, and not the effect of the quantities they purchase. Finally, some of the over-time price variation could also be problematic if stores run promotions during periods of seasonal demand, and households tend to watch more ads on TV during those periods, say because manufacturers buy more (or better) advertising at the national-level at those times. Therefore, a credible IV strategy requires a) isolating *within*-brand, *within*-store, *over*-time price variation that moves quantities around, *after* controlling for seasonality, and b) checking whether ad consumption responds to the movement in quantities induced by that variation. The remainder of this section first documents that such price variation exists in the data and then discusses the IV results.

Figure 6 plots time series of prices for stores 8 and 9, the two most frequently visited stores in the data. 85% of the households in our data make at least one purchase from these stores. Figure **??** shows extensive over-time variation in prices. Appendix C shows analogous plots for the full set of stores. As a way of summarizing this, for each brand, we compute the coefficient of variation (CV) in weekly prices for each store. Figure 7 presents box-plots of the CVs for all the stores for each of the 11 brands. On average, the within-store standard deviation in prices is seen to be of the order of 10-15% of the mean price for the brand at that store.

[14]This viewpoint has parallels in the applied econometric literature. For example, Angrist and Krueger (1991) estimate the effect of schooling on earnings using quarter of birth as an instrument for years of education. The typical expectation is that those of higher ability will find schooling easier and will obtain more schooling to signal their ability. Thus, *a priori* we may expect that OLS estimates of earnings on years of schooling are upward biased because of omitted unobserved ability that is positively correlated with earnings and schooling. Alternatively, it may be possible that there is no signaling, or that some individuals with higher earning potential drop out of school earlier to pursue their own endeavors. On instrumenting for years of schooling, Angrist and Kreuger find the IV coefficient to be positive and slightly larger than the OLS estimate in several specifications, indicating if anything that OLS is slightly biased downward. Ashenfelter and Krueger (1994) report large increases in the estimated returns to education after correcting for measurement error in self-reported years of schooling and omitted variables bias. Card (2001) notes that it is common to find that IV estimates of the returns to schooling tend to exceed their OLS counterparts.

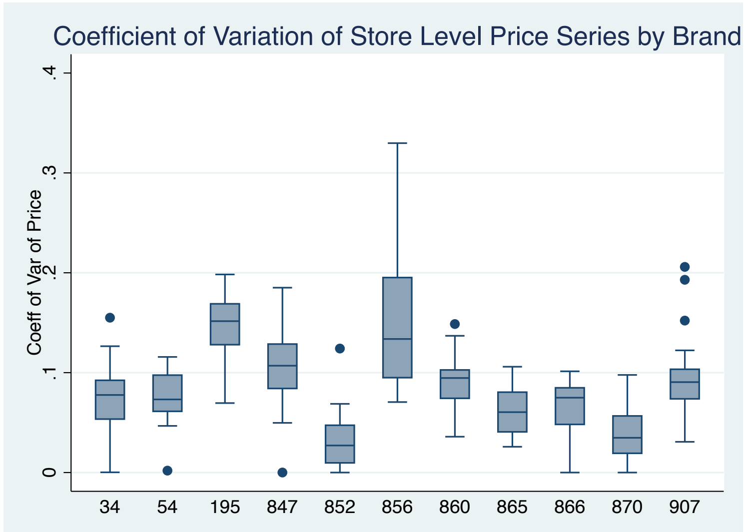
Figure 6: Price Variation over Time at Stores 8 and 9



Note: The figure plots the time-series of prices of the 11 brands at stores 8 & 9, the two most frequently visited in the data (85% of the households in the data make at least one purchase from these stores). Each line corresponds to a separate brand.

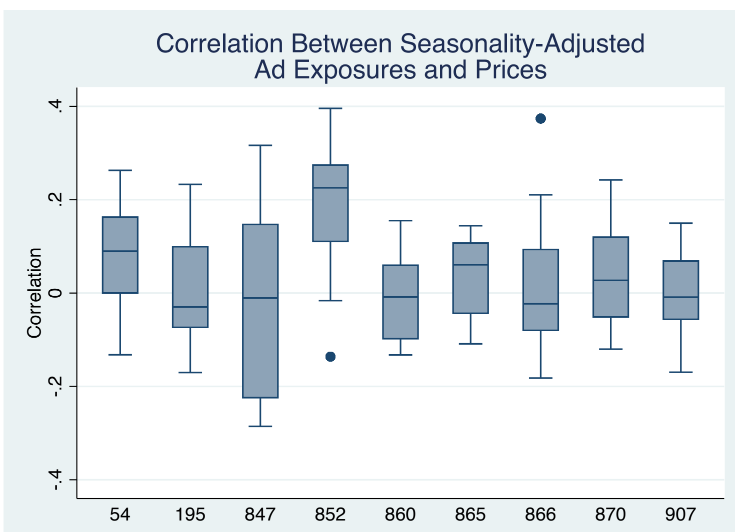
We now check whether there is evidence in the data of coordination of retail prices and TV advertising. To eyeball this, Figures 9 and 10 illustrate the covariation in ad exposures and prices at stores 8 & 9. For each brand, we aggregate all the ad exposures to consumers in the data each week and plot a time series of these aggregate exposures against the price of that brand at the two stores. We see evidence for independent movement in prices and advertising. Recall that we will control for seasonality in our regressions using a full set of week fixed effects, so what is relevant is whether there is a systematic relationship between advertising and store level prices *after* controlling for such seasonality. Figure 8 reports box-plots of seasonality-adjusted correlations between ad exposures and store-level prices, separately for each brand. In order to make the seasonality-adjustment, we regress the weekly price series pooled across all brands and stores on a set of store, brand, and week FEs. Similarly, we regress weekly ad exposures pooled across brands on a set of brand and week FEs. Finally, we calculate the correlation between the residuals from these two regressions for each store and brand. For all brands, we fail to reject the null that there is systematic correlation in the level of ad exposures and the prices faced by consumers at the stores in the sample. After controlling for seasonality, there does not seem to be evidence of coordination between retail prices and more intensive advertising on TV, which would be problematic for our identifications strategy. In this regard, the data is consistent with what we know about the institutional context.

Figure 7: Coefficient of Variation of Store Level Price Series by Brand



Note: For each brand along the row, the vertical box-plot summarizes the distribution across stores of the store-specific coefficients of variation of weekly prices for that brand.

Figure 8: Correlation Between Seasonality-Adjusted Ad Exposures and Prices



Note: For each brand along the row, the vertical box-plot summarizes the distribution across stores of the correlation between aggregate ad exposures and store-specific weekly prices for that brand, after purging seasonality.

Figure 9: Time Series Plots of Ad Exposures and Prices by Brand at Store 8

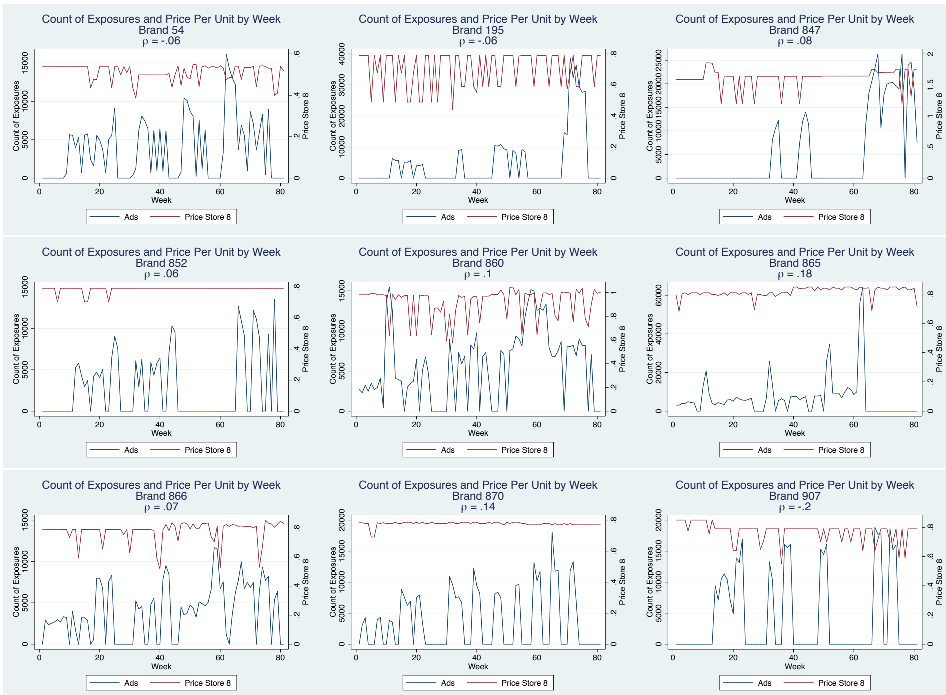
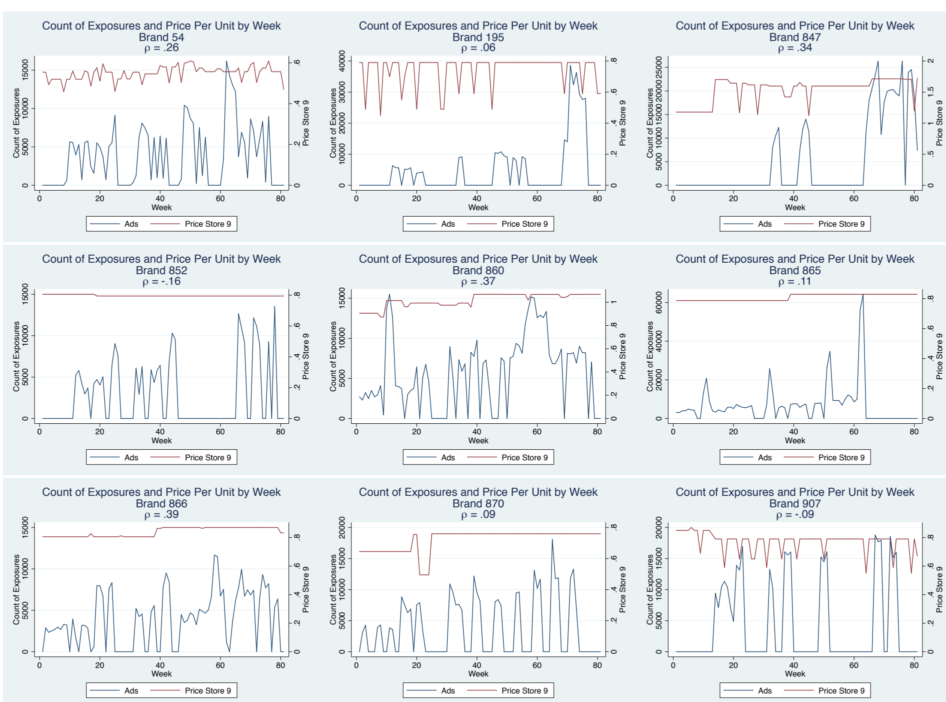


Figure 10: Time Series Plots of Ad Exposures and Prices by Brand at Store 9



Results We estimate Equation (3) via 2SLS. Since we include brand and week fixed effects, our estimates are identified off the within-brand variation induced in quantities by prices after purging seasonality. We present two different specifications that instrument for past quantity purchased using past prices for the product faced by the consumer. The first specification uses as instruments prices of all the brands at stores 8 and 9, the two most frequently visited retailers. As most households make at least one purchase from these two stores, these prices are salient to most. The second specification uses as instruments a household-specific price series we construct for each brand. To do this, we weight the store-level price series for each brand by the percentage of purchases made by that household at each store over the duration of the data. Since this gives more weight to prices at the stores a consumer is more likely to visit, these instruments are likely to have more power in explaining quantities. For this latter specification, there may be a concern that unobservable consumer tastes may shift both the store a household chooses to visit and ad-skipping behavior. Therefore, we strengthen controls for unobserved heterogeneity more by including a full set of household-specific fixed effects in one specification, and a fixed effect for each household's most frequently visited store in another. Both IV specifications use prices for all brands to instrument for quantity, so the models are over-identified.

Tables 7 and 8 present the IV results. The first-stage is reported in Appendix D. Looking at Tables 7 and 8, we see that the IV estimates of past purchases of the product on current ad consumption are positive and significant. We also see that a consistent feature across specifications is that the IV coefficient is larger than the FE coefficient. To emphasize this, in both tables, we report the ratio of the estimated IV coefficient to that of the FE coefficient reported in column 1, Table (5). The IV coefficient is seen to be 15 to 90 times larger. The direction of movement in the IV coefficient relative to the FE model is robust across specifications and controls. Our broad take-way from these results is that FE appears downward biased.

Tables 7 and 8 also report on the first-stage F statistics for each of the 2SLS regressions. The first-stage F statistics are around 6 for the specifications in Table 7, and around 9 for the specifications in Table 8. Given the lower values, we may be worried about a finite sample bias in these estimates arising from a weak instruments problem. In general, we worry about weak instruments when the instruments' correlation with the endogenous regressor is low and when there are many over-identifying restrictions. In such a case, 2SLS is known to be biased in the direction of OLS (Stock, Wright and Yogo, 2002). Given this, we also report instrumental variables specifications using the limited information maximum likelihood (LIML) estimator. LIML is median-unbiased for over-identified models with weak instruments and provides more robust inference in the presence of weak IVs. Comparing the 2SLS and LIML estimates in Tables 7 and 8, we find that the LIML point estimate is comparable to or larger than the 2SLS estimate. The LIML point estimates are relatively larger than the 2SLS estimates in the specifications that have

Table 7: IV Regressions of Ad Consumption on Cumulative Quantity: Prices for All Brands at Top 2 Stores as Instruments

	(1) Percent Ad Watched Conditional on Exposure	(2) Percent Ad Watched Conditional on Exposure	(3) Percent Ad Watched Conditional on Exposure
2SLS			
$\hat{\tilde{Q}}_{ijt,14}$	0.052***	0.052***	0.050***
	(0.017)	(0.017)	(0.017)
LIML			
$\hat{\tilde{Q}}_{ijt,14}$	0.145**	0.147**	0.116**
	(0.065)	(0.066)	(0.049)
Week FE	Y	Y	Y
Brand FE	Y	Y	Y
Store FE	N	Y	N
HH FE	N	N	Y
$\frac{Coeff_{IV}}{Coeff_{FE}}$	54	54	52
First Stage F	6.10	6.13	5.79
Anderson-Rubin $\chi^2(24)$	379.63	380.64	348.61
Anderson-Rubin p-Value	0.000	0.000	0.000
Observations	1,436,400	1,436,400	1,436,400
Marginal Effect of an Additional Purchase of Average Quantity on Expected Percentage Watched			
	5.2%	5.2%	5.0%
Standard errors in parentheses			
*** p<0.01, ** p<0.05, * p<0.1			

Note: Regressions estimated at the household-brand-day-exposure level. The dependent variable Percent Ad Watched records the percentage of the exposure that was watched and ranges between 0 and 1. Regressions estimated conditional on an exposure. $\tilde{Q}_{ijt,14}$ records the cumulative package volume household i purchased of brand j in the 14 days preceding day t, normalized by the average daily purchase volume, $\bar{q}_t$ = 2,015 equivalent units. Past consumption instrumented using the average prices of all brands over the 14 days preceding day t in the two stores with the highest market share. The model is over-identified with 1 endogenous regressor and 24 excluded price instruments. Robust standard errors clustered at the household level.

a smaller first stage F statistic. In specifications where the first-stage F is large (e.g., Table 8), the LIML estimate is close to the 2SLS estimate. These comparisons suggest some correction for weak-IV bias in 2SLS by the LIML estimator. The broader take-away is that weak IV bias does not drive the fact that the instrumental variables estimates are found to be larger than the fixed effects estimates across specifications.

Finally, we also report the Anderson-Rubin (1949) statistic for a test of $\theta_1 = 0$ (i.e, the null hypothesis that past quantity has no effect on ad consumption) joint with a null that the overidentifying restrictions are valid. The AR test is robust to weak IVs. In most specifications we continue to find evidence consistent with complementarities.

Taken together, these instrumental variables results tell a consistent story – namely, that the FE estimates appear to be downward biased and thus present a conservative test of the model.

Table 8: IV Regressions of Ad Consumption on Cumulative Quantity: HH Price for All Brands as Instruments

	(1) Percent Ad Watched Conditional on Exposure	(2) Percent Ad Watched Conditional on Exposure	(3) Percent Ad Watched Conditional on Exposure
2SLS			
$\hat{\tilde{Q}}_{ijt,14}$	0.092***	0.039**	0.014
	(0.023)	(0.019)	(0.015)
LIML			
$\hat{\tilde{Q}}_{ijt,14}$	0.105***	0.034*	0.012
	(0.026)	(0.019)	(0.015)
Week FE	Y	Y	Y
Brand FE	Y	Y	Y
Store FE	N	Y	N
HH FE	N	N	Y
$\frac{Coeff_{IV}}{Coeff_{FE}}$	96	41	15
First Stage F	9.41	9.48	8.84
Anderson-Rubin $\chi^2(12)$	31.20	14.18	8.09
Anderson-Rubin p-Value	0.002	0.289	0.778
Observations	1,436,400	1,436,400	1,436,400
Marginal Effect of an Additional Purchase of Average Quantity on Expected Percentage Watched			
	9.2%	3.9%	1.4%

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Regressions estimated at the household-brand-day-exposure level. The dependent variable Percent Ad Watched records the percentage of the exposure that was watched and ranges between 0 and 1. Regressions estimated conditional on an exposure. $\tilde{Q}_{ijt,14}$ records the cumulative package volume household i purchased of brand j in the 14 days preceding day t, normalized by the average daily purchase volume, $\bar{q_t}$ = 2,015 equivalent units. Past purchase instrumented using the average prices of all brands over the 14 days preceding day t. The model is over-identified with 1 endogenous regressor and 12 excluded price instruments. Robust standard errors clustered at the household level.

Additional Cuts Appendix E reports on additional specifications on the dual side of the model, i.e., the effect of advertising on quantities purchased. Past advertising is found to shift quantities purchased. Appendix F then explores whether complementarities operate at the category level, as opposed to the brand-level as modeled above. The results are consistent with brand level complementarities, and not category level complementarities.

Summary To summarize these results, we find that households tend to watch more advertising on the margin when they have purchased more in the past, and they tend to buy more when they have been exposed to more advertising in the past. These patterns do not seem to be driven by correlated tastes that drive both product purchase and ad viewing behavior, and provide support for a model of complementarities.

In the following section we present a model of demand for goods and advertising that allows for, but does not impose, complementarities between purchases and advertising. The model will allow us to relate the demand for products and advertising to a well defined utility maximization problem; to implement more efficient joint estimation of the simultaneous equations system defining advertising and product demand; to control for correlated unobserved heterogeneity and allow for stock effects; and to evaluate counterfactual scenarios that explore the response in purchases and welfare to changes in advertising while taking into account the co-dependence between purchases and ads.

4 A Model with Complementarities

We describe an empirical model of consumption of goods and of advertisements. The consumption of goods is central to most utility maximization problems and is described in the next section. The consumption of advertisements is novel, however, and requires some attention. Unlike the case of regular consumption of goods, exposure to advertisements is not always deliberate. Advertising seeps into consumers' daily activities, often interrupting their favorite television programs and internet browsing. While consumers may not be able to choose which commercials they are exposed to on TV, they have agency over whether to watch a whole commercial or to skip it. This is the decision that we model: We investigate whether consumers skip advertisements conditional on advertising exposure. Before we describe the main parts of the model, we present two key assumptions.

Assumption 1: Sequential Choices We treat the good and advertising consumption decisions as related but separate. While we allow the decisions to be interdependent (through complementarities or common factors that affect both actions), the decisions of which goods to buy and how much of each advertisement to watch are separated in time and location and unlikely to be simultaneously made. It is more likely that consumers make these decisions sequentially during the day. Hence, we use a sequential model of daily decisions as depicted in Figure 11. We assume that within a given day, the purchase decision of goods takes place before the decision of advertising consumption. Our data suggest this assumption is reasonable. Our advertising data includes a timestamp which indicates the exact second of each exposure. Looking at Figure 12 below, we see that 70% of ad exposures occur after 5 PM. Unfortunately, we do not observe a time stamp in the purchase data to test the time of purchase of products, but it is not unreasonable to assume that many in-store purchases occur during the day.

Figure 11: Timing of Consumption Decisions

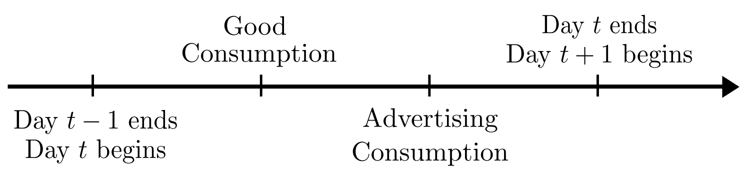
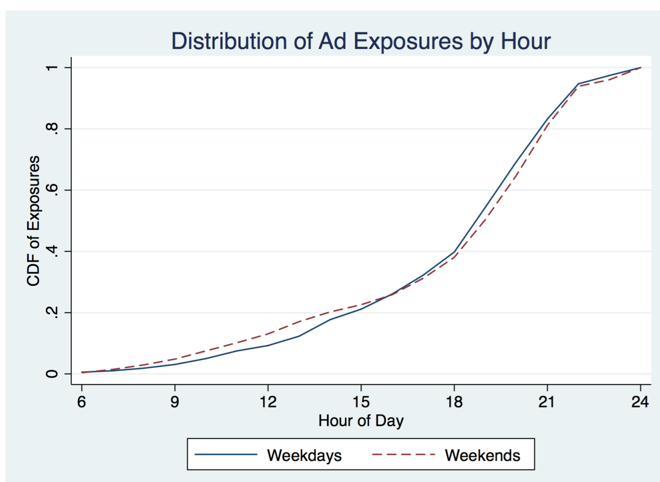


Figure 12: Ad Exposures by Time of Day



Assumption 2: Myopic Behavior We also assume that consumers make purchase and ad-skipping decisions myopically, without formally incorporating the effect of their purchase decisions on subsequent ad consumption utility, and without incorporating the effect of their ad-skipping decisions on subsequent product purchase utility. Consumers face a number of decision problems during the course of a day and it is hard to imagine that they would, for example, deliberately decide to buy more of a good in order to enjoy its advertisements more later in the same day. Further, it seems implausible that ad-skipping is driven significantly by the fact that a consumer may anticipate that seeing more or less of an ad may

change his utility from *future* product consumption. While possible to incorporate at substantially higher computational cost, to us, entertaining such degree of forward-looking decision-making in this context seems unrealistic.

Finally, in the structural model of ad-skipping below, we model only a binary variable of the consumer's action of seeing the ad fully versus skipping it towards another channel or activity. Thus, on the ad-side, we model only the skip or not decision, and not the continuous decision of *how much* of an ad to watch. We think this modeling choice is more consistent with the actual trade-off consumers make and the key preference over advertisements is revealed in the decision to skip the ad or not.[15] We do, however, think that the amount of ad time watched by consumers may be correlated with purchase outcomes. For example, if a commercial is informative or contains a catchy jingle, then even if the ad was not watched to completion (was "skipped"), a consumer who saw 95% of the ad may be more likely to make a purchase than a consumer who saw only 5% of the ad, i.e., an ad may have an effect on purchases even if it was skipped. To capture this, we use the continuous variation in ad consumption when modeling purchase decisions.

In the following, we utilize a discrete-continuous model of demand following Wales and Woodland (1983) and Kim, Allenby and Rossi (2002). The discrete-continuous model has the advantage of flexibly handling corner solutions and continuous purchase quantities across brands while retaining a clear link to a direct utility function (please see Chintagunta and Nair 2011 for a review). Later, when we present our counterfactuals, we show that explicitly handling the quantity decisions in the model has a material effect on our welfare and profit assessments. Within this set-up, we use Bhat (2005) and Lee, Kim and Allenby's (2013) parametrization of utility, which has been documented to fit scanner panel data well. We relax the models above by introducing the possibility of advertising acting as a complement to the advertised product.

4.1 Consumption Utility of Goods

On day t, an agent decides whether and how much to consume of each of $J+1$ goods by maximizing a direct utility function conditional on past advertising exposures:

$$\max_{\boldsymbol{x}_t \geq 0} \quad U^G\left(\boldsymbol{x}_t | \boldsymbol{A}_{t-1}\right) \quad s.t. \; \boldsymbol{p}_t^G \cdot \boldsymbol{x}_t \leq E_t^G \tag{4}$$

where $\boldsymbol{x}_t$ is a vector of product quantities $x_{0t}...x_{Jt}$, $\boldsymbol{p}_t^G$ is a vector of prices $p_{0t}^G...p_{Jt}^G$, and E_t^G is the consumer's total expenditure. Here, $\boldsymbol{A}_{t-1}$ is a vector of the stock of ad consumption over the preceding 14 days for each product, $A_{1,t-1}...A_{J,t-1}$. Some amount of the outside good ($j=0$) is always consumed,

[15] In Appendix G, we show that our results are not sensitive to local changes to how we define an ad as "skipped". Our results remain unaffected if we define an ad as "skipped" if the fraction viewed is < 1, 0.95, 0.9, 0.8 or 0.75. Similarly, the descriptive results from our main specification in Column 1 of Table 5 do not change qualitatively if we instead employ the binary skip or watch indicator as the dependent variable.

and its price is normalized to one. The total direct utility from consumption is divided into two sub-utility functions as shown below, where U_0^G captures the utility from consuming the outside good and U_1^G captures the direct utility from consumption of the remaining goods as well as complementarity effects with advertising exposure:

$$U^G\left(x_{0t}...x_{Jt}|\,\boldsymbol{A}_{t-1}\right)=U_0^G\left(\cdot\right)+U_1^G\left(\cdot\right) \tag{5}$$

where

$$U_0^G\left(x_{0t}\right)=e^{\gamma_0}x_{0t}\times e^{\mu\varepsilon_{0t}^G} \tag{6}$$

$$U_1^G\left(x_{1t}...x_{Jt}|\,\boldsymbol{A}_{t-1}\right)=\sum_{j=1}^{J}\exp\left[\gamma_j+\beta\log\left(1+A_{jt-1}\right)\right]\log\left(1+x_{jt}\right)\times\mathrm{e}^{\mu\varepsilon_{jt}^G} \tag{7}$$

Above, $\varepsilon_{0t}^G...\,\varepsilon_{Jt}^G$ are i.i.d. stochastic shocks that are known to the consumer but not to the econometrician, assumed to be T1EV distributed and μ is a scale parameter. We introduce μ so the expected demands implied by the problem are well defined (discussed in more detail below). The utility is quasilinear which is consistent with the fact that the category in question makes up a small percentage of household income, and income effects are not first-order.[16] The utility function is additively separable across brands, reflecting the assumption that the brands are substitutes. However, non-separability in utility between ads and quantities implies complementarities between consumption and advertising of a given brand are allowed. The implied marginal utility of consumption is,

$$\frac{\partial U^G}{\partial x_{jt}}=\frac{\exp\left[\gamma_j+\beta\log\left(1+A_{jt-1}\right)\right]}{\left(1+x_{jt}\right)}\times e^{\mu\varepsilon_{jt}^G} \tag{8}$$

The marginal utility of consuming x_{jt} depends on the quantity purchase x_{jt} and on the quantity of advertising the individual consumed for that good in the past 14 days, A_{jt-1}. The complementary effect of advertising on the marginal utility from consumption of good j is given by,

$$\frac{\partial}{\partial A_{jt-1}}\frac{\partial U^G}{\partial x_{jt}}=\frac{\beta\exp\left[\gamma_j+\beta\log\left(1+A_{jt-1}\right)\right]}{\left(1+x_{jt}\right)\left(1+A_{jt-1}\right)}\times e^{\mu\varepsilon_{jt}^G} \tag{9}$$

An attractive feature of this model is that the sign of β determines whether advertising complements or substitutes the consumption of good j.

The agent maximizes her utility subject to the budget constraint in the domain of positive quantities. The Lagrangian for the utility maximization problem is,

$$\max_{\boldsymbol{x}_t\geq 0,\lambda_t\geq 0}\mathbb{L}^G\left(x_t,\lambda\right)=U^G\left(\boldsymbol{x}_t|\,\boldsymbol{A}_{t-1}\right)+\lambda_t\left(E_t^G-\sum_j\boldsymbol{p}_{jt}\cdot\boldsymbol{x}_{jt}\right) \tag{10}$$

which can be solved by the Karush-Kuhn-Tucker (KKT) optimality conditions. One goal of the demand model is to accommodate the mixed distribution of zeros (corner solutions) and positive quantities observed in the data. A generic observation comprises no-purchase of some brands, and positive quantities

[16] Equation (5) is quasilinear because by dividing through by $e^{\gamma_0+\mu\epsilon_{0t}^G}$, a monotone transformation of $U^G\left(.\right)$, we can write $U\left(x_{0t}...x_{Jt}|\,\boldsymbol{A}_{t-1}\right)=x_{0t}+U\left(x_{1t}...x_{Jt}|\,\boldsymbol{A}_{t-1}\right)$.

of the others by a household on a given day. Without loss of generality, let the first K goods be consumed, and the consumption of the rest be zero. Then the KKT conditions for this observation imply that,

$$\frac{\partial \mathbb{L}^G}{\partial x_{jt}} = 0, \quad x_{jt}^* > 0, \, j \in (0,..,K) \tag{11}$$

$$\frac{\partial \mathbb{L}^G}{\partial x_{jt}} \leq 0, \quad x_{jt}^* = 0, \, j \in (K+1,..,J) \tag{12}$$

$$x_{0t}, x_{jt}, \lambda_t \geq 0, \quad j \in (0,..,J) \tag{13}$$

The problem above can be simplified by solving the optimality conditions with respect to λ_t and taking logarithms. Following the standard procedure of differencing with respect to the outside good (see for e.g., Kim, Allenby and Rossi 2002), it follows that at the optimum,

$$\mathcal{V}_{0t} - \mathcal{V}_{jt} = \varepsilon_{jt} - \varepsilon_{0t}, \, x_{jt}^* > 0, \, j \in (1,..,K) \tag{14}$$

$$\mathcal{V}_{0t} - \mathcal{V}_{jt} \geq \varepsilon_{jt} - \varepsilon_{0t}, \, x_{jt}^* = 0, \, j \in (K+1,..,J) \tag{15}$$

where,

$$\mathcal{V}_{0t} = \tfrac{\gamma_0}{\mu} \tag{16}$$

$$\mathcal{V}_{jt} = \tfrac{1}{\mu}\left[\gamma_j + \beta \log\left(1 + A_{jt-1}\right) + \log\left(1 + x_{jt}\right) - \log\left(p_{jt}\right)\right] \tag{17}$$

In the remainder of this section, we derive in sequence the expected demand; the expected consumer surplus, and the joint probabilities of purchase and density of quantities implied by this model. In our empirical analysis, we will use the expected demand and consumer surplus to calculate expected profits and welfare from counterfactual targeting scenarios; and the implied joint density of purchase/quantities to form a maximum likelihood estimator of the consumer's utility parameters.

4.1.1 Expected Demand

To simplify notation, we define $\eta_{jt} \equiv \varepsilon_{jt}^G - \varepsilon_{0t}^G$. η_{jt} is the difference between two T1EV random variables and has a logistic distribution with location 0 and scale 1. Equation (15) of the KKT conditions implies that if the consumer buys any units of brand j, it has to be that marginal utility for the initial unit of good j is higher than that of the outside good, i.e.,

$$\mathcal{V}_{jt}\big|_{x_{jt}=0} + \varepsilon_{jt}^G > \mathcal{V}_{0t} - \varepsilon_{0t}^G \tag{18}$$

$$\iff \quad \eta_{jt} > \underline{\eta}\left(A_{jt-1}, p_{jt}\right) \equiv \quad -\tfrac{1}{\mu}\left[\gamma_j - \gamma_0 + \beta \log\left(1 + A_{jt-1}\right) - \log\left(p_{jt}\right)\right] \tag{19}$$

We can combine this with Equation (14) to characterize the quantity of good j at time t bought by a given consumer conditional on demand shock η_{jt}:

$$x_{jt}^*\left(A_{jt-1}, p_{jt}, \eta_{jt}\right) = \begin{cases} \frac{\kappa_{jt}(A_{jt-1})}{p_{jt}} \exp\left(\mu \eta_{jt}\right) - 1, & \eta_{jt} \geq \underline{\eta}\left(A_{jt-1}, p_{jt}\right) \\ 0 & \eta_{jt} < \underline{\eta}\left(A_{jt-1}, p_{jt}\right) \end{cases} \tag{20}$$

where $\kappa_{jt}(A_{jt-1}) = \exp[\gamma_j - \gamma_0 + \beta \log(1 + A_{jt-1})]$.

Integrating over the demand shock η_{jt} yields the expected demand for the consumer,

$$\mathbb{E}_\eta[x^*(A_{jt-1}, p_{jt}, \eta_{jt})] = \int_{\underline{\eta}(A_{jt-1}, p_{jt})}^{\infty} \left(\frac{\kappa_{jt}(A_{jt-1})}{p_{jt}} \exp(\mu\eta_{jt}) - 1\right) d\mathcal{F}_\eta(\eta_{jt}) \tag{21}$$

where $\mathcal{F}_\eta$ is the c.d.f. of a standard logistic distribution. Given the properties of the logistic distribution, a necessary condition for the first moment of quantity to be finite is that $\mu < 1$.

4.1.2 Consumer Welfare

Given the quasilinear utility, the consumer surplus is an exact representation of consumer welfare (representing both equivalent or compensating variation; see, for eg., Varian 1992, chap. 10). We calculate the consumer surplus by integrating the demand function in (20) from the observed price up to the reservation price at which the consumers' demand becomes zero. From equation (20), we can see that the reservation price $\bar{\mathfrak{p}}(A_{jt-1}, \eta_{jt})$ at which demand $x^*(A_{jt-1}, \bar{\mathfrak{p}}_{jt}, \eta_{jt}) = 0$ is,

$$\bar{\mathfrak{p}}(A_{jt-1}, \eta_{jt}) = \kappa_{jt}(A_{jt-1}) \exp(\mu\eta_{jt}) \tag{22}$$

which defines the maximum price the consumer is willing to pay for a unit of good j at time t. Therefore, the surplus to the consumer from buying inside good j conditional on the demand shock η_{jt} is,

$$CS_{jt}(A_{jt-1}, p_{jt}, \eta_{jt}) = \mathbb{I}[\bar{\mathfrak{p}}(A_{jt-1}, \eta_{jt}) > p_{jt}] \int_{p_{jt}}^{\bar{\mathfrak{p}}(A_{jt-1}, \eta_{jt})} \left[\kappa_{jt}(A_{jt-1}) \frac{\exp(\mu\eta_{jt})}{p} - 1\right] dp \tag{23}$$

where p_{jt} is the per-unit price of good j available to consumer i at time t. The integral above sums the consumer surplus from the price observed in the data (p_{jt}) up to the reservation price $\bar{\mathfrak{p}}(A_{jt-1}, \eta_{jt})$. The indicator function $\mathbb{I}[\bar{\mathfrak{p}}(A_{jt-1}, \eta_{jt}) > p_{jt}]$ ensures that consumer surplus is only calculated when a positive quantity of the good is bought. The expected consumer surplus (unconditional on η_{jt}) is,

$$\mathcal{E}CS_{jt} = \mathbb{E}_\eta[CS_{jt}(A_{jt-1}, p_{jt}, \eta_{jt})] = \int_{-\infty}^{\infty} CS_{jt}(A_{jt-1}, p_{jt}, \eta_{jt})\, d\mathcal{F}_\eta(\eta_{jt}) \tag{24}$$

Calculating the integral in (23) with respect to prices, and substituting into (24), the total expected consumer surplus is,

$$\mathcal{E}CS_{jt} = \int_{\widetilde{\eta}_{jt}(A_{jt-1}, p_{jt})}^{\infty} \left\{p_{jt} + \kappa_{jt}(A_{jt-1}) \exp(\mu\eta_{jt}) \left[\mu\eta_{jt} + \log\left(\frac{\kappa_{jt}(A_{jt-1})}{p_{jt}}\right) - 1\right]\right\} d\mathcal{F}_\eta(\eta_{jt}) \tag{25}$$

where the lower bound for the demand shocks, $\widetilde{\eta}_{jt}(A_{jt-1}, p_{jt}) = \frac{1}{\mu} \log \frac{p_{jt}}{\kappa_{jt}(A_{jt-1})}$ arises from the indicator function in (23).

Neither the expected demand in equation (21) nor the expected consumer surplus in equation (25) can be computed analytically; in our counterfactuals, we compute these integrals by simulation.

4.1.3 Probability of Purchase and Density of Quantities

We close this section by presenting the probabilities of purchase and the associated density of quantities implied by the model, which we use later to construct the likelihood. To do this, we refer back to the KKT conditions for a generic observation in the data for which the first K goods are purchased, and the consumption of the rest is zero. For the goods not purchased, equation (15) implies that their unobserved stochastic utility components relative to the outside good, η_{jt}, cannot be larger than $\mathcal{V}_{0t} - \mathcal{V}_{jt}$. The corresponding probability that these goods are not purchased can be obtained by integrating the density of η_{jt} over the truncated support consistent with no-purchase. For the goods purchased, the model implies optimal quantities purchased are determined by trading off the marginal utility from consumption of the brand with that of the outside good. Equation (14) captures this tradeoff. Since the quantity purchased depends on η_{jt}, the distribution of η_{jt} induces a distribution on quantities, which can be derived by change-of-variables calculus using equation (14). Assuming $\varepsilon_{0t}^{G}, ..., \varepsilon_{Jt}^{G}$ follows a Type-1 extreme-value distribution, the mixed discrete-continuous density of a generic observation for an agent at time t conditional on a vector of prices and past advertising becomes,

$$l\left(x_{0t}^{*}, ..., x_{Kt}^{*}, 0, ..., 0 \middle| \boldsymbol{p}_{t}, \boldsymbol{A}_{t-1}\right) = K! \times \frac{\Pi_{j=0}^{K} \exp\left(\mathcal{V}_{jt}\right)}{\left[\Sigma_{j=0}^{J} \exp\left(\mathcal{V}_{jt}\right)\right]^{K+1}} \times \left|\mathcal{J}_{t}\right| \tag{26}$$

where $V_{.t}$ are as defined in Equations (16 and 17) and$|\mathcal{J}_{t}|$ is the determinant of the Jacobian induced by the nonlinear change in variables transformation from the density of the error terms to the density of the purchased quantities. The Jacobian $\mathcal{J}_{t}$ is diagonal with elements,

$$\left[\mathcal{J}_{t}\right]_{l,k} = \frac{\partial\left(\mathcal{V}_{0t} - \mathcal{V}_{lt}\right)}{\partial x_{kt}^{*}} = \mathbf{1}\left(l = k\right) \frac{1}{\mu\left(1 + x_{kt}^{*}\right)}, \; l, k = 1..K \tag{27}$$

and its determinant is $\left[\Pi_{k=1}^{K} f_{kt}\right]$, where $f_{kt} = \frac{1}{\mu\left(1+x_{kt}^{*}\right)}$, $k = 1, .., K$ (see Bhat 2005; Chintagunta and Nair 2011; or Lee, Kim and Allenby's 2013 for derivations in related setups).

4.2 Consumption Utility of Advertising

We now introduce the consumer's advertising consumption decision. We assume that advertising exposures are independent and take place sequentially within a day. An observation is a household-day-exposure combination. At advertising exposure s for brand j in day t, the consumer decides $a_{sjt} \in \{0, 1\}$ where a zero means the advertisement is skipped (opt for the outside option) and a one means that the consumer watches the advertisement in its entirety. We view the ad-skip decision as the outcome of a time-allocation problem. We assume that an ad is τ seconds long, so not skipping it consumes τ seconds out of T total seconds allocated by the consumer for watching the show airing at a given occasion. Skipping the ad takes the consumer to the outside option (e.g., watching the show) of which some continuous quantity is always consumed. The number of seconds allocated to the outside option is denoted w_{st}.

Conditional on an advertising exposure, the consumer's problem is given by,

$$\max_{w_{st}\in\mathbb{R}^{+},a_{sjt}\in(0,1)} \quad U^{A}\left(w_{st},a_{sjt}\right|X_{jt}) \quad s.t. \;\; w_{st}+\tau a_{sjt}\leq T \tag{28}$$

where X_{jt} is a stock of past product consumption in the preceding 14 days (including the purchase earlier on day t, if applicable). Analogous to the purchase-side model, we assume the direct utility $U^{A}(.)$ is,

$$U^{A}\left(w_{st},a_{sjt}\right|X_{jt}) = U_{0}^{A}\left(w_{st}\right)+U_{1}^{A}\left(a_{sjt}\right|X_{jt}) \tag{29}$$

where,

$$\begin{aligned} U_{0}^{A}\left(w_{st}\right) &= e^{\alpha_{0}}w_{st}\times e^{\varepsilon_{s0t}^{A}} \\ U_{1}^{A}\left(a_{sjt}\right|X_{jt}) &= a_{sjt}\exp\left[\alpha_{j}+\theta\log\left(1+X_{jt}\right)\right]\times \mathrm{e}^{\varepsilon_{sjt}^{A}} \end{aligned}$$

and $\varepsilon_{s0t}^{A},\varepsilon_{sjt}^{A}$ are IID Type-1 extreme value errors that shift the value from skipping versus seeing the ads. To solve for the choices induced by this program, recall, a_{st} is binary. If the consumer decides to skip exposure s, $a_{sjt}=0$. Substituting for a_{sjt} into the time-constraint in Equation (28), $w_{st}+\tau\times 0=T$, so $w_{st}=T$ when $a_{sjt}=0$. We can now obtain the conditional indirect utility from skipping by evaluating the direct utility (28) at $w_{st}=T$ and $a_{sjt}=0$,

$$\begin{aligned} \mathcal{V}^{A}\left(a_{sjt}=0|X_{jt}\right) &= U_{0}^{A}\left(w_{st}=T\right)+U_{1}^{A}\left(a_{sjt}=0\right|X_{jt}) \\ &= e^{\alpha_{0}}T\times e^{\varepsilon_{s0t}^{A}} \end{aligned}$$

If the consumer decides to watch ad exposure s, $a_{sjt}=1$. Substituting again for a_{sjt} into the time-constraint in Equation (28), $w_{st}+\tau\times 1=T$, so $w_{st}=T-\tau$ when $a_{sjt}=1$. We can obtain the corresponding conditional indirect utility from not skipping by evaluating the direct utility (28) at $w_{st}=T-\tau$ and $a_{sjt}=1$,

$$\begin{aligned} \mathcal{V}^{A}\left(a_{sjt}=1|X_{jt}\right) &= U_{0}^{A}\left(w_{st}=T-\tau\right)+U_{1}^{A}\left(a_{sjt}=1\right|X_{jt}) \\ &= e^{\alpha_{0}}\left(T-\tau\right)\times e^{\varepsilon_{s0t}^{A}}+\exp\left[\alpha_{j}+\theta\log\left(1+X_{jt}\right)\right]\times \mathrm{e}^{\varepsilon_{sjt}^{A}} \end{aligned}$$

The consumer will not skip the ad if $\mathcal{V}^{A}\left(a_{sjt}=1\right|X_{jt})>\mathcal{V}^{A}\left(a_{sjt}=0\right|X_{jt})$. Taking logarithms on both sides of this inequality, and letting $\mathcal{W}_{sjt}=\alpha_{j}+\theta\log\left(1+X_{jt}\right)$, and $\mathcal{W}_{s0t}=\alpha_{0}+\log\left(\tau\right)$, the consumer will watch the advertisement if,

$$\mathcal{W}_{sjt}-\mathcal{W}_{s0t}\geq\varepsilon_{s0t}^{A}-\varepsilon_{sjt}^{A} \tag{30}$$

The implied probability of watching an advertisement is a logit,[17]

$$Pr\left(a_{sjt}=1\right)=\frac{\exp\left(\mathcal{W}_{sjt}\right)}{\exp\left(\mathcal{W}_{s0t}\right)+\exp\left(\mathcal{W}_{sjt}\right)} \tag{31}$$

[17] τ is not identified separately from the intercept and is absorbed into α_{0}.

Evidence of complementarities for the agent can be found if,

$$\frac{\partial}{\partial X_{jt}}\left[U_1^A\left(a_{sjt}=1\right|X_{jt}\right)-U_1^A\left(a_{sjt}=0\right|X_{jt}\right)\Big]=\frac{\theta}{1+X_{jt}}\exp\left[\alpha_j+\theta\log\left(1+X_{jt}\right)\right].\mathrm{e}^{\varepsilon_{sjt}^A}\geq 0 \qquad (32)$$

Again, an attractive feature of the specification is that the interaction parameter θ captures the direction of complementarity/substitution between advertising and consumption. A positive value of θ would indicate that the consumption of good j increases the utility of its advertisements.

4.3 Maximum Likelihood

Purchase Likelihood The mixed discrete-continuous density of a purchase observation for an individual was presented in Equation (26). Because purchase incidence is infrequent at the daily level, we estimate the model conditional on purchase of at least one of the inside goods. This keeps the empirical model from losing precision from having to fit a large number of zeros. This also means we only identify parameters from the quantity and brand-choice variation, and not from the buy vs. not-buy decision. Let B_t indicate the purchase of an inside good s.t.,

$$B_t=\begin{cases}1 \text{ if } x_{jt}>0 \text{ for at least one } j\in\{1,..,J\}\\ 0 \text{ otherwise}\end{cases}$$

To handle the selection conditional on purchase, we derive the corresponding conditional purchase likelihood as the unconditional likelihood divided by the probability of purchase,

$$\mathcal{L}_t^G=l\left(x_{0t}^*,...,x_{Kt}^*,0,...,0\right|B_t=1\big)=\frac{l\left(x_{0t}^*,...,x_{Kt}^*,0,...,0\right)}{1-Pr(B_t=0)}$$

where,

$$Pr(B_t=0)=\frac{\exp(\mathcal{V}_{0t})}{\sum_{j=0}^J\exp(\mathcal{V}_{jt})}$$

Ad Consumption Likelihood While we infer ad consumption purely from the consumer's decision to skip or watch an ad, for predictive purposes it is useful to have a statistical model for the number of ad exposures and the brand of each exposure. Let z_t denote the total number of ads a consumer is exposed to on day t, b_{1t},...,b_{zt} indicate the brand of each ad exposure, and a_{1t},...,a_{zt} indicate the consumer's binary decision to skip or watch those exposures. The joint likelihood of observing the ad exposures, the brand content of the exposures, and the ad-skip decisions is,

$$\begin{aligned}\mathcal{L}_t^A&=Pr\left(z_t,b_{1t},...,b_{zt},a_{1t}^*,...,a_{zt}^*\right|X_t\big)\\&=\underbrace{Pr\left(a_{1t}^*,...,a_{zt}^*\right|z_t,b_{1t},...,b_{zt},X_t\big)}_{\text{Logit}}\underbrace{Pr\left(b_{1t},...,b_{zt}\right|z_t,X_t\big)}_{\text{Ind. Draws from pmf}}\underbrace{Pr(z_t|X_t)}_{\text{Poisson}}\end{aligned}$$

We choose statistical distributions to fit the observed exposures and ad-brand content across agents in the data. To accommodate the large number of zero exposures, we assume z_t follows a zero-inflated Poisson distribution with probability mass function:

$$Pr(z_t = h) = \begin{cases} \pi + (1-\pi)e^{-\lambda}, & h = 0 \\ (1-\pi)\frac{\lambda^h e^{-\lambda}}{h!}, & h \geq 1 \end{cases} \tag{33}$$

where λ and π are parameters to be estimated. Next, let $b_{st} \in (1, .., J)$ denote the brand corresponding to exposure s on day t. We assume b_{st} follows a multinomial distribution with parameters ϕ_j s.t. $\sum_{j=1}^{J} \phi_j = 1$. Assuming the probabilities are independent across exposures, the likelihood of viewing the observed distribution of ad-brand content, conditional on z_t exposures is,

$$Pr\left(b_{1t}, ..., b_{zt} \middle| z_t, X_t\right) = \prod_{s=1}^{z_t} \prod_{j=1}^{J} \phi_j^{1(b_{st}=j)} \tag{34}$$

Finally, under the assumption that $\boldsymbol{\varepsilon}^A$ is T1EV, the problem of maximizing utility through the consumption of advertisements is a discrete choice model much in the spirit of McFadden (1974) and subsequent classical applications. The last component of the conditional likelihood of the consumer's choices is a logit likelihood,

$$l\left(a_{0t}^*, ..., a_{zt}^* \middle| z_t, b_{1t}, ..., b_{zt}, X_{t-1}\right) = \prod_{s=1}^{z_t} \left(\frac{\exp\left(\mathcal{W}_{sjt}\right)}{\exp\left(\mathcal{W}_{s0t}\right) + \exp\left(\mathcal{W}_{sjt}\right)} \right)^{a_{sjt}} \left(\frac{\exp\left(\mathcal{W}_{s0t}\right)}{\exp\left(\mathcal{W}_{s0t}\right) + \exp\left(\mathcal{W}_{sjt}\right)} \right)^{1-a_{sjt}} \tag{35}$$

To reiterate, our ad-utility parameters $(\boldsymbol{\alpha}, \theta)$ are identified purely off the conditional skip/no-skip decision (equation 35), and not from the observed distribution of exposures and ad-brand content across agents in the data. We fit a distribution to exposures and brand-content purely for prediction purposes.

Deriving the Joint Likelihood Putting the two pieces together, we arrive at the joint likelihood of observing the purchase and advertising data given the model parameters,

$$\mathcal{L}_t = \mathcal{L}_t^G \times \mathcal{L}_t^A$$

We allow for a random effects specification of heterogeneity. To reflect this, we now introduce the index i for agent. Collect the key parameters of interest in a vector $\Theta_i = (\boldsymbol{\gamma}_i, \beta_i, \boldsymbol{\alpha}_i, \theta_i)$. We assume that

$$\Theta_i \sim MVN(\bar{\Theta}, \Sigma)$$

We allow the key parameters pinning down the complementarities in the purchase and ad consumption models β and θ, to covary with each other. This allows us to capture the fact that households whose consumption utility is sensitive to their level of ad consumption may also derive ad utility that is sensitive to the level of product consumption. Such households may experience a feed-back loop whereby viewing a large number of ads leads them to buy a lot of the product which in turn leads them to view more ads. In addition, we allow for the fact that households who tend to have a high preference for brand j may also derive higher utility from advertisements for brand j by allowing γ_j and α_j to be correlated. Accordingly, Σ is specified to be a block-diagonal matrix with the exception of non-zero covariance terms that we estimate on the $\gamma_j \times \alpha_j$ and $\beta \times \theta$ off-diagonals. We estimate this joint model via maximum simulated likelihood using 1,000 draws of Θ_i per household to integrate over the implied random effects distribution. The mixed discrete-continuous density implied by the purchase model and the joint covariance with the advertising model combined with the large panel duration per household (557 days) makes the likelihood function complicated to maximize. We use the efficient SNOPT solver with a likelihood tolerance of $10e-6$ to facilitate the maximization. On an AMD 64-bit Unix server equipped with a 24-core Intel Xeon X5650 chip running at 2.67GHz and parallelized across 12 workers, the model takes roughly 7 days to maximize.

5 Estimation Results

The estimates for the parameters of the product and ad consumption equations are presented in Table 9. The estimates of the remaining auxiliary parameters (λ, π, ϕ_j) are included in a table in Appendix H. Looking first at the estimates for the product consumption model in Table 9, the $\bar{\gamma}$ parameters reflect brand-level purchase incidence in the data, such that brands with a more negative coefficient are those that are purchased less frequently. The negative values reflect the large share of no-purchase at the daily level. The relatively large magnitude of the σ^{γ} parameters indicates significant heterogeneity in purchase frequency across households. The complementarity coefficient on advertising $\bar{\beta}$ is estimated to be positive and the relatively small magnitude of σ^{β} implies a low probability of having a negative β. Turning to the estimates for the advertising consumption model, recall that brands 1 and 6 do not advertise in our data, so we do not estimate parameters for these brands. The intercept parameters $\bar{\alpha}$ pin down ad-skip rates, so ads for brands with a higher intercept correspond to those observed to be skipped less frequently. The relatively large magnitude of the σ^{α} parameters shows that there is extensive heterogeneity in ad-skip rates across households. The mean complementarity parameter $\bar{\theta}$ in the ad consumption model is estimated to be positive albeit not statistically significant. The standard deviation σ^{θ} however, is relatively large and significant, and it identifies a distribution of complementarity (and substitution) effects across households. To understand the economic significance of the complementarities encapsulated in these estimates, we will report several targaeting counterfactuals below. The estimated correlation between β and θ is captured by ρ and is estimated to be positive but small. Finally, though the scale parameter μ is theoretically identified, we were unable to pin it down across a range of specifications we tried. We calibrate $\mu = \frac{1}{2}$ and estimate the remaining parameters of the model fixing μ at this value. We re-estimated the model

Table 9: Parameter Estimates for the Joint Model Estimated on the Full Sample

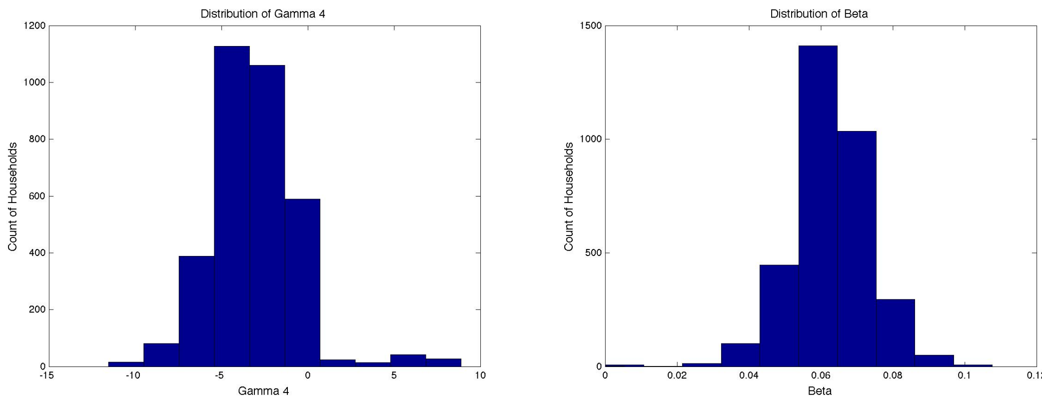
Brand	$\bar{\gamma}_{ij}$	σ_{ij}^{γ}	$\bar{\alpha}_{ij}$	σ_{ij}^{α}
1	-2.9555	0.3624	-	-
	(0.0218)	(0.0286)	-	-
2	-2.9506	0.4092	3.1207	0.3823
	(0.0208)	(0.0241)	(0.0175)	(0.0304)
3	-0.8893	3.4289	3.3273	0.2131
	(0.0229)	(0.0151)	(0.0178)	(0.0449)
4	-2.7595	3.2278	3.2790	0.2908
	(0.0384)	(0.0378)	(0.0162)	(0.0307)
5	-2.4964	0.3820	2.7269	0.5520
	(0.0181)	(0.0219)	(0.0177)	(0.0275)
6	-1.8154	2.5924	-	-
	(0.0377)	(0.0321)	-	-
7	-2.3140	0.1944	2.7002	0.6312
	(0.0162)	(0.0241)	(0.0140)	(0.0195)
8	-4.9805	4.3027	2.8935	0.6482
	(0.0571)	(0.0416)	(0.0155)	(0.0215)
9	-2.6011	0.1427	2.5248	0.7409
	(0.0174)	(0.0346)	(0.0160)	(0.0206)
10	-2.6809	0.1969	2.6825	0.5932
	(0.0205)	(0.0430)	(0.0161)	(0.0227)
11	-2.1150	3.8769	2.9105	0.4888
	(0.0296)	(0.0227)	(0.0174)	(0.0256)

$\bar{\beta}$	σ^{β}	$\bar{\theta}$	σ^{θ}	ρ
0.0631	0.0138	0.0012	0.0332	0.0004
(0.0067)	(0.0107)	(0.0040)	(0.0059)	(0.0003)

LL = -1,699,456.84
No. of Households = 4,221
No. of Observations = 2,325,771

Notes: Table reports results from joint maximum likelihood estimation of the purchase and advertising model across households. Random effects allowed on $\Theta_i = (\boldsymbol{\gamma}_i, \beta_i, \boldsymbol{\alpha}_i, \theta_i)$ as $\Theta_i \sim MVN(\bar{\Theta}, \Sigma)$, where, Σ is a block-diagonal matrix with the exception of non-zero covariance terms on the $\gamma_j \times \alpha_j$ and $\beta \times \theta$ off-diagonals. The parameter ρ measures the covariance between β and θ. The covariances between γ_j and α_j are estimated but not reported. Model estimated via maximum simulated likelihood using 1,000 draws of Θ_i per household to integrate over the implied random effects distribution. The scale μ is fixed at 0.5.

Figure 13: Household Specific Estimates of γ_4 and β



at various values of $\mu \in (-1, 1)$ and found the fit of the model to the data to be roughly the same.

After estimating the demand system, we use the estimated mean $\hat{\bar{\Theta}}$ and variance $\hat{\Sigma}$ parameters to characterize the distribution of household-specific tastes. We apply Bayes Rule to calculate an estimate of each household's expected preference parameters conditional on the household's purchase history and the estimated distribution of the population's preference parameters as,

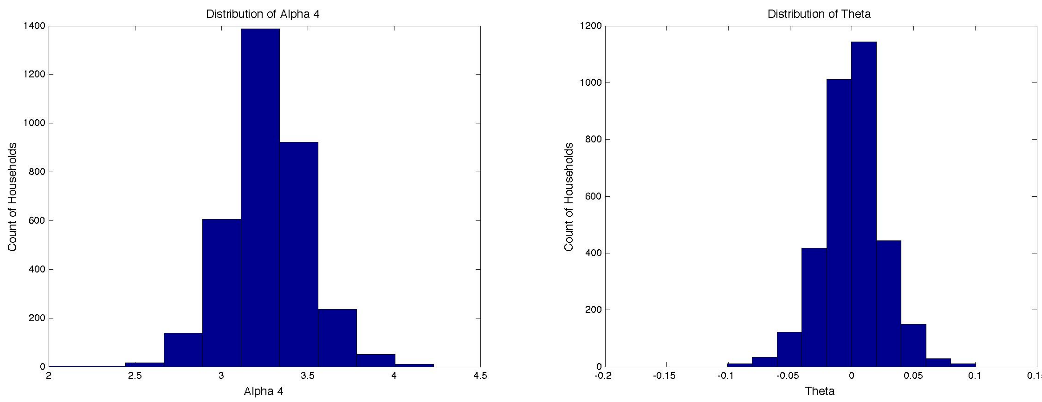
$$\hat{\bar{\Theta}}_i = \int \Theta_i \frac{\mathcal{L}_i(\Theta_i|\hat{\bar{\Theta}}, \hat{\Sigma})}{\mathcal{L}_i(\hat{\bar{\Theta}}, \hat{\Sigma})} d\Phi(\Theta_i|\hat{\bar{\Theta}}, \hat{\Sigma}) \tag{36}$$

We use this "approximate Bayesian" approach (Allenby and Rossi 1999; Revelet and Train 2001; Chintagunta et al. 2005) to recover household level parameter estimates that can be used to measure heterogenous treatment effects.

5.1 Model Simulations

Assessing intuition regarding the implication of estimates in a nonlinear model with random effects like our model above is best done with simulation. We would like to document what the estimates predict about changes in demand in response to changes in advertising and to illustrate endogenous non-compliance with treatment. We first forward-simulate daily advertising and purchase outcomes for the last quarter of the data using our estimated household-specific coefficients at the observed prices and level of ad exposure. We then simulate advertising and purchase outcomes increasing the number of ad exposures for each brand separately by one standard deviation, holding everything else fixed. For illustration, we discuss in detail below the output of this simulation for brand 4. To aid in the interpretation of the simulations, we include plots of the household specific preference parameters for brand 4 in Figures 13 and 14. Table 11 at the end of this section summarizes the analogous results for all the advertised brands in the data.

Figure 14: Household Specific Estimates of α_4 and θ



Consider brand 4. We simulate the increase in exposures for the brand as follows. We first calculate the total number of exposures each household saw for brand 4 in the last quarter of 2011. The mean number of exposures over the three months is 24. We then increase each household's number of ad exposures by the standard deviation of the number of brand 4 Q4-2011 ad exposures across households. This turns out to be 28 exposures. We allocate the additional 28 exposures evenly across the days in which each household was observed to view an ad for brand 4 in the data. Our simulation only includes the 2,751 households who viewed at least one advertisement for brand 4 during these 3 months.

Endogenous Non-Compliance: Ad Skipping Figure 15 shows the model-predicted household skip rates (fraction of ad exposures that are skipped) at the observed level of exposures. The median household skip rate is 0.0351 which is similar to the overall percent of ads that are observed to be skipped in the data. These skip rates can be thought of as a measure of the uptake of treatment. All households are "offered" the same treatment of an additional 28 advertisements, but households vary in the extent to which they are treated, and, importantly, this variation in treatment is an endogenous outcome of the model.

Figure 16 plots a histogram across households of the increase in realized total ad consumption resulting from the increased level of exposures. The realized total ad consumption for a household is calculated by computing the model-predicted proportion of each of the 28 additional exposures that is consumed, and then adding up these proportions across the 28 incremental exposures. This plot shows that the extensive heterogeneity in skip rates documented in Figure 15 results in differences in the level of treatment across households. Although all households were shown an additional 28 ad exposures, the average increase in ad consumption ranges from 26.1 to 27.9 ads across households. This reflects the differential uptake of treatment across households.

Figure 15: Distribution of Simulated Household Ad Skip Rates of Brand 4 Exposures

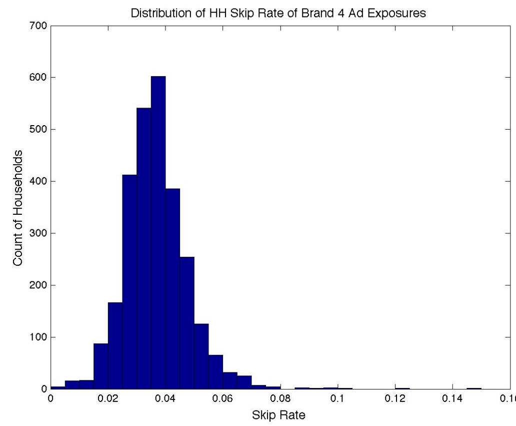
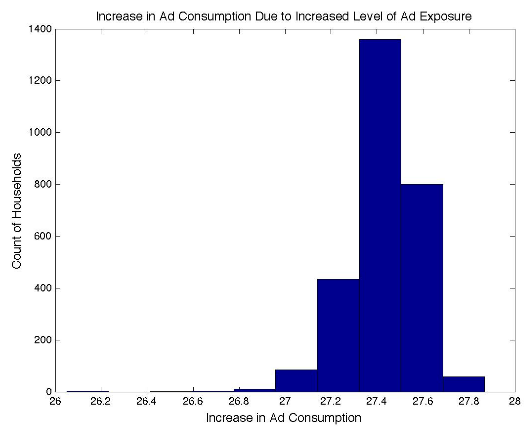


Figure 16: Increase in Consumption of Brand 4 Ads Resulting from an Increase in Number of Ad Exposures

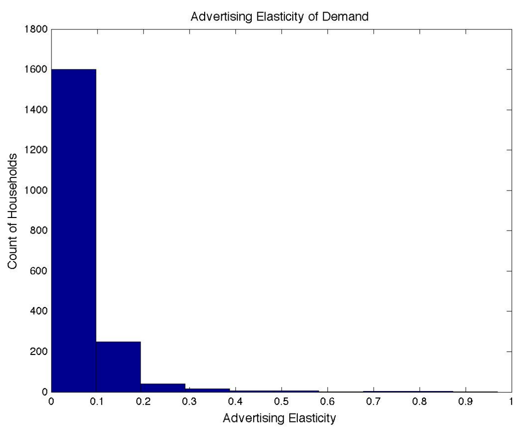


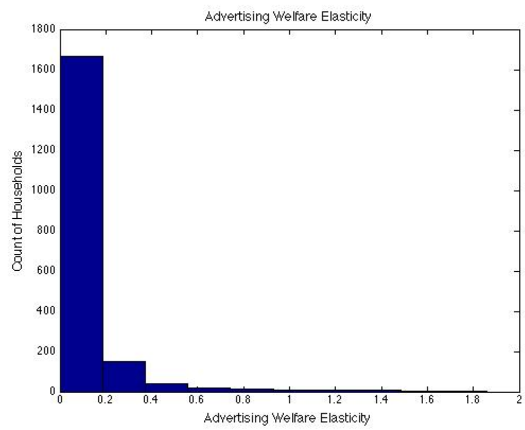
Endogenous Non-Compliance: Measuring Treatment Effects We now discuss the implications of the differential uptake of advertising on purchase quantity and welfare. In order to measure the responsiveness of demand and welfare to changes in advertising, we calculate advertising elasticities. Looking at equation (37), we calculate the advertising elasticity of demand for each household as the household's percent increase in total predicted purchase quantity of brand 4 over the three months in the simulation, divided by the corresponding percent increase in predicted ad consumption. For each household, we calculate welfare in a given day using equation (24). We sum this consumer surplus across the last three months of the data to obtain the total surplus from consumption. We report a "welfare elasticity" calculated as the percent increase in total consumer surplus over the three months in the simulation divided by the percent increase in ad consumption.

$$\eta_{ij}^{q} = \frac{\Delta Q_{ij}/Q_{ij}}{\Delta A_{ij}/A_{ij}}, \eta_{ij}^{u} = \frac{\Delta \mathcal{C}S_i/\mathcal{C}S_i}{\Delta A_{ij}/A_{ij}} \tag{37}$$

Appendix I describes the simulation steps in greater detail. To the extent these measure tabulate in outcomes over a quarter in response to changes in the consumed advertising stock, these elasticities capture medium to long-run effects of advertising. Figure 17 shows the distributions of model-predicted demand and welfare elasticities with respect to advertising across households. The median advertising elasticity of demand is 0.0561 and the median advertising elasticity of welfare is 0.0639.[18] There is significant heterogeneity across households.

Figure 17: Advertising Elasticities





The model implies that households self-select into receiving more or less treatment by choosing to skip ads. Logically, we should also expect that those households who watch more ads are the ones that show larger purchase quantity and welfare changes. We assess informally whether households who are predicted by the model to consume more incremental ad-time also see larger increases in these outcome variables of interest, thereby assessing whether endogenous non-compliance with advertising matters at the estimated parameters. We regress the increase in predicted quantity and utility on the model-predicted incremental ad time watched. The results are in Table 10. In both, the effect of incremental ad time is positive − i.e., those who consume more ad-time see larger increases in purchase outcomes and welfare. In effect, this illustrates that the change in ad consumption can explain the cross-sectional change in purchase quantity and welfare, but the change in ad exposures cannot (because in this exercise, all households are subject to the same change in ad exposures). This is important for firms who may be considering the profitability of a targeted ad campaign because the efficacy of the campaign will depend on the compliance of the sub-population that is targeted. A positive take-away is that if the advertiser can observe individual-level ad consumption and not just ad exposure, the firm can identify the subset of households whose purchases

[18]Across brands, we estimate average advertising elasticities of demand ranging from 0.02 to 0.06 (Table 11).

Table 10: Regression of Increase in Purchase Outcomes on Increase in Ad Consumption

	(1)	(2)
	Δ Purchase Quantity$_i$	Δ Consumer Surplus$_i$
$\Delta \mathrm{A}_i$	293.15***	4,004.69***
	(88.33)	(1,274.10)
Constant	-7,984.63***	-109,114.60***
	(2,422.91)	(34,949.98)
Observations	2,751	2,751
Mean Dep Var.	56.75	737.03

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Regression estimated at the household level. The independent variable $\Delta \mathrm{A}_i$ records each household's model-predicted increase in ad consumption between the two simulations. The dependent variables record the model-predicted increase in purchase quantity and consumer surplus between the two simulations.

and welfare will likely change in response to the campaign. We explore this further in the targeting counterfactuals below.

Table 11: Summary of Simulation Results by Brand

Brand	Δ Ad Exposures	Δ Ad Consumption	Demand Elasticity	Welfare Elasticity
2	8	7.8190	0.0234	0.0252
		(0.0687)	(0.0227)	(0.0281)
3	25	24.5314	0.0328	0.0366
		(0.1080)	(0.0451)	(0.1079)
4	28	27.4511	0.0561	0.0639
		(0.1467)	(0.2572)	(1.9588)
5	9	8.7017	0.0265	0.0274
		(0.1374)	(0.0219)	(0.0260)
7	11	10.6414	0.0297	0.0318
		(0.2058)	(0.0258)	(0.0305)
8	0	-	-	-
		-	-	-
9	12	11.5512	0.0299	0.0322
		(0.3053)	(0.0280)	(0.0325)
10	6	5.8014	0.0157	0.0158
		(0.1106)	(0.0134)	(0.0179)
11	10	9.7280	0.0207	0.0221
		(0.1181)	(0.0649)	(0.2663)

Note: Table reports the median and standard deviation of outcomes across households. Each simulation includes the set of households who were exposed to at least one ad for the focal brand. One outlier household dropped in the simulation for brand 4.

6 Counterfactuals

Our counterfactuals are motivated by the future of TV ad markets, in which new digital streaming and set-top box technology has made televisions individually addressable (via their IP-address or unique set-top box identifiers). Compared to online advertising markets, TV ad markets currently provide only limited targeting ability, allowing advertisers to buy ad spots on TV shows whose attractiveness in reaching desired audiences can be assessed on the basis of only limited, aggregated audience data split by coarse demographics (age and gender). The coming addressability of TV in the future will imply that advertisers will be able to target ads to individual consumers *directly*, rather than to the shows they watch. Further, increasingly, as TV viewing data gets merged with data on the corresponding product purchases of the viewers, advertisers will be able to target audiences more finely on the basis of their preferences for products (as revealed in their historical purchase data), rather than on the basis of coarse demographics like age and gender. More broadly, in this addressable ecosystem advertisers could contemplate targeting ads to consumers who are less likely to skip them and who are more likely to respond favorably to advertising consumption by increasing product demand. Non-skipped ads may be favored by the TV network all things equal, as they reduce the chance that targeted consumers move away from the show being viewed or from viewing TV to other entertainment options. Further, as product purchases are tracked, it may be feasible to target both advertising and prices (say via targeted discounts or coupons) to pinpointed audiences, and to track subsequent redemptions and purchases. The structural model endogenizes product purchase and ad consumption behavior in a fully specified setup with heterogeneous preferences and is useful to assess how demand, consumer welfare and profitability may change in such an addressable market.

We use our model to assess a variety of targeting scenarios. First, we consider how demand, welfare, and profits would change if an advertiser could target ads to consumers who are less likely to skip them (i.e., essentially those with $\frac{\partial U}{\partial A} > 0$), where we assess expected skip-rates using our model and estimated parameters. The goal is to assess whether ad-viewing activity information provides targeting value for firms and whether that benefits or hurts consumers. To the extent that it is possible for advertisers to assess skip-rates of consumers by leveraging only TV viewing data, and to the extent that TV networks may prefer that ads are targeted to the subset of their viewers who are less likely to skip them, this kind of targeting policy is likely to be increasingly used by firms in the future. Complementarities in demand for products and advertisements imply that those who like the ads are also more likely to prefer the advertisers' products; hence, this ad-targeting strategy benefits the advertiser by indirectly picking consumers who are favorably disposed to its product.

Intuitively, this targeting policy could be improved if the advertiser also incorporated product preference information. Because some who do not skip ads may have bought the product anyway even in the absence of ad exposures, it is possible that targeting on ad skip-rates alone allocates advertising to infra-marginal consumers and misses marginal consumers. Since we have estimated product and ad preferences, we can use the full model to assess outcomes if the advertiser were able to target consumers for whom the marginal ad induces a positive change in purchase behavior ($\frac{\partial^2 U}{\partial A \partial Q} > 0$). The scenario calculates outcomes as if the advertiser were able to optimize its allocations under full information of consumers' preferences.

Finally, complementarities imply that prices play a stronger role in moderating advertising effects in our model compared to other setups for handling advertising. Because the propensity to consume advertising is mediated by how much of the product is consumed, low prices can induce higher ad consumption on the margin via their effect on product quantity demanded. The increased ad consumption in turn affects product demand, and so on, resulting in potential feedback and large returns to low prices. To assess this, we simulate a third counterfactual in which we allow the advertiser to target both individual prices and ads to consumers, essentially implementing first-degree price and advertising discrimination. For each, we compute demand, consumer welfare and profitability across the consumers in the data.

Our results are meant to illustrate the importance of considering demand-side complementarities and the value of endogenizing the decision to consume advertising in assessing these targeting scenarios. An important caveat to the counterfactuals is that we do not incorporate competitive price and advertising response in reaction to the improved price and advertising targeting by the focal advertiser. Thus, our results are not meant to speak to equilibrium outcomes in a market with improved addressability and targeting. Doing so would require specifying a supply-side model of price and advertising competition, which is beyond the scope of the current demand-side analysis.

6.1 Simulation Setup and Procedure

We implement our counterfactuals using the 2011 Fall season in our data as a benchmark. We first forward-simulate daily advertising and purchase outcomes for the 106 days starting Sept 17, 2011 through Dec 31, 2011 for all households in our data using our estimated household-specific coefficients at the observed prices and level of ad exposure. Then we ask how outcomes look under various counterfactual ad and price targeting scenarios compared to this benchmark. Rather than predicting how many customers may watch TV on a given day (and hence can potentially be targeted with ad exposures), we implement all our counterfactuals holding the number and the timing of total exposures fixed at the allocation observed in Fall 2011 in the data. Thus, we implicitly hold the sequence of "opportunities" to deliver exposures, as well as the total number of possible exposures fixed under all counterfactual comparisons. For example, suppose we observe b_{jt} exposures by advertiser j on day $t = 1, .., 106$ in Fall 2011 in the data. In our counterfactuals, we hold b_{jt} fixed for each t and vary how the b_{jt} exposures are allocated across different sets of consumers. Thus, the ad-side control variable for the firm in all our counterfactuals is a set of indicators $\left\{\tilde{b}_{ijt}; i = 1, .., N\right\}$ such that $\tilde{b}_{ijt} = 1$ if consumer i gets allocated ads on day t, and 0 otherwise, and such that $\sum_{i=1}^{N} \tilde{b}_{ijt} = b_{jt} \, \forall t$. In counterfactuals in which price targeting is also considered, the control variable for the firm also includes $\tilde{p}_{ijt}$, an individual price to offer consumer i for purchase of product j in period t. We treat firms as myopic, by not incorporating into their price and advertising decisions that current controls can affect future profits through their effects on the consumer's consumption stocks. This implies that our static assessments will understate the potential impact on profits from targeting. Assuming constant marginal costs, the firm's expected profit from consumer i on day t is,

$$\pi_{ijt} = (p_{ijt} - c) \times \mathbb{E}\left[x^*_{ijt}\right] - c_b b_{ijt} \tag{38}$$

where $\mathbb{E}\left[x^*_{ijt}\right]$ is as defined in equation (21) and c_b is the cost of an ad exposure. The total profit at time t is,

$$\pi_{jt} = \sum_{i=1}^{N} (p_{ijt} - c) \times \mathbb{E}\left[x^*_{ijt}\right] - c_b b_{jt} \tag{39}$$

In counterfactuals invoking only ad targeting, we allocate ads to consumers on the basis of their expected ad-skip probability, where the probability is implied by the ad-skip model in equation (31). In counterfactuals involving only targeted pricing, we solve for the optimal prices for each consumer by maximizing the profit function in equation (38) with respect to prices. Because the optimization with respect to $\mathbb{E}\left[x^*_{ijt}\right]$ is complex, we utilize an approximation to the optimal price by solving a first-order Taylor-series approximation to the first-order conditions implied by the pricing problem. Appendix J provides exact details. In counterfactuals in which we consider ad- and price-targeting, we use the full ad and product joint demand model, computing the expected profit to the firm for each consumer when targeting an optimal price, taking into account that the consumer can choose to skip or watch the ad with some probability implied by the model. Appendix K presents exact details of the reallocation algorithms along with pseudo-code for the steps implemented. In all counterfactuals, we fix the marginal cost c by assuming the manufacturer earns a margin 30% on prices observed in the data.

While the full profits include the cost to the brands of buying ad exposures, in the results reported below we only compute operating profits, $\sum_{i=1}^{N}(p_{ijt}-c)\times\mathbb{E}\left[x^*_{ijt}\right]$ under the status-quo and other counterfactual scenarios, and do not include the advertising cost, cb_{jt}. In the absence of data on ad-prices and a formal model of the ad-supply market, we are unable to back out how the costs of exposures (c_b) change under the current and counterfactual scenarios. The measures we report should therefore be interpreted as the gross value of the future addressable advertising technology to the channel as a whole, which will then be split between advertisers and TV networks according to bargaining and negotiation power.

6.2 Results: Profits

The results for profits assessments under the various counterfactual scenarios are presented for two cases: one, where the product prices are held fixed at the values in the data (Table 12), and two, for the case where firms are allowed to price their products at the household level (Table 13). The benchmark referred to as "Status Quo" refers to profits under the observed ad allocation in the data; the "Ad viewing" case refers to the ad-allocation in which households are targeted on the basis of their expected skip-rate, and the "Full Information" case refers to the situation where the firm can use the full model of product and ad-response to target as described above.

Table 12: Normalized Profits with Fixed Prices

					Brand					Weighted
Targeting Method	2	3	4	5	7	8	9	10	11	Average
Status Quo: Data	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Targeting: Ad Viewing	99.9	101.4	106.3	99.8	99.1	100.5	100.4	100.0	99.7	101.3
Targeting: Full Information	103.2	108.3	112.1	103.6	101.1	105.9	102.8	101.0	107.7	108.5

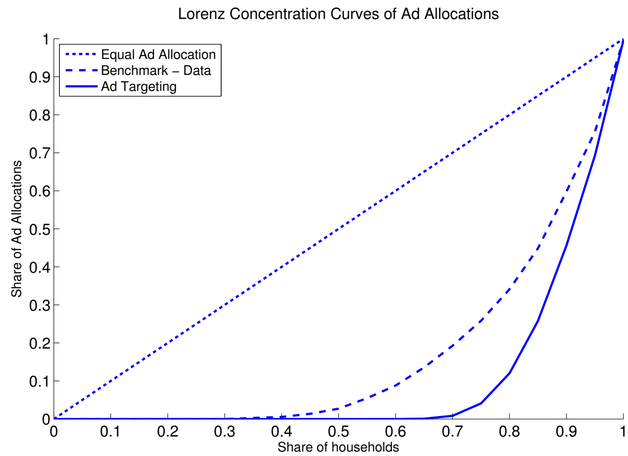
Table 13: Normalized Profits with Targeted Prices

Targeting Method	2	3	4	5	Brand 7	8	9	10	11	Weighted Average
Status Quo: Data	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Targeting: Ad Viewing	99.1	101.4	107.0	100.5	100.8	100.5	101.9	101.3	99.6	101.4
Targeting: Full Information	100.7	108.3	112.1	102.4	101.3	105.9	100.5	100.5	108.1	108.6

We first discuss the case with no price discrimination. Looking at Table 12, we see that when prices are fixed, being able to target households based on their expected ad-viewing ("Targeting: Ad Viewing" case) increases profits by 1.3% on average relative to the benchmark across brands. Targeting on the basis of full information increases profits by 8.5%. Taken together, these figures mean that on average, targeting on the basis of ad-viewing makes up about 15.3% − (101.3-100)/(108.5-100) − of potential profits attainable through full targeting, revealing that ad-viewing behavior is likely to be a relevant piece of targeting information for firms. There is considerable heterogeneity across brands in these effects. For instance, for brand 4, these benefits are high − yielding 52.1% − (106.3-100)/(112.1-100) − of the revenues generated by full targeting. However, for some brands, targeting on ad-viewing alone is inferior to the TV-show-based ad allocation policy firms are currently implementing, primarily because the estimated correlations between brand preferences and ad preferences in the population are not significantly positive in those cases. Full-information based targeting is seen to benefit all brands relative to the status-quo. Looking at Table 13, when prices are allowed to be adjusted to the HH-level, we find that the relative uptake for Full-Information based targeting versus ad-viewing based targeting parallels the results when prices are held fixed. In particular, targeting on the basis of ad-viewing makes up about 16% − (101.4-100)/(108.6-100) − of potential profits attainable through full targeting. Thus, it seems that there is not a quantitatively and economically significant interaction between price discrimination and its impact on the returns to ad-targeting.

To provide some insight into what is driving these results, we present a more detailed analysis of the outcomes for brand 4. Figure 18 presents the Lorenz curves associated with the benchmark and the ad-viewing targeting cases for brand 4. To interpret this figure, note that higher levels of concentration of ad exposures across households are associated with lower Lorenz curves (more to the right). The straight line denotes the case where ads are equally distributed across consumers. The benchmark case (dashed line) denotes the concentration of ads distributed to consumers with regular ad targeting technologies. Finally, the targeting case (solid line) denotes the concentration of ad allocations to consumers by ad-viewing behavior. It is clear that ad targeting concentrates ad exposures on a relatively small portion of households over time. In particular, before the targeting policy is considered, about 20% of households

Figure 18: Lorenz Curves of Benchmark and Ad-viewing Targeting

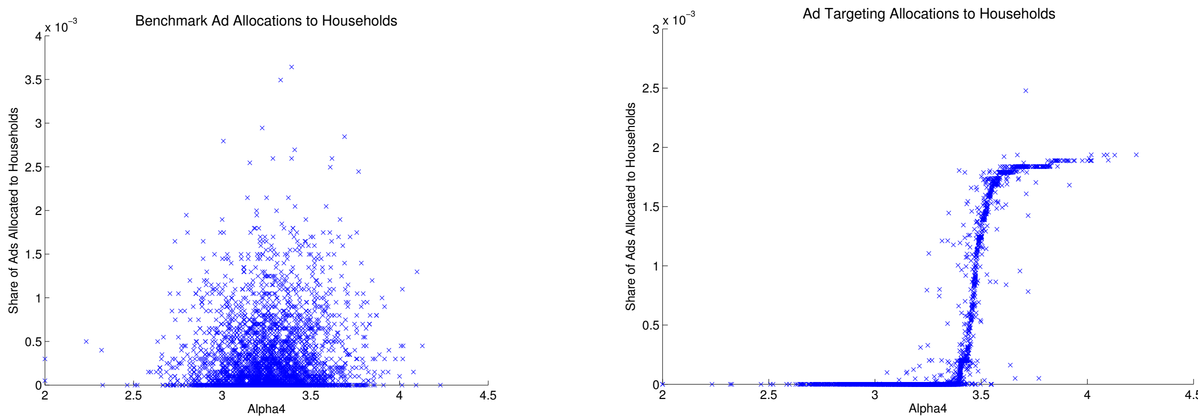


received roughly 65% of ad exposures. With ad-viewing based targeting, the percentage of ad exposures to the households increases to about 90% of the total number of ad exposures. To obtain intuition for which subset of households the new policies allocate advertising to, we check the correlation of the share of advertising allocated to a household under the counterfactual with its estimated parameters. We find that allocation is mainly driven by parameter α_{i4}, the intercept in the ad-skip model in equation (31). The correlation between ad watching behavior of a household and its α_{i4} and γ_{i4} parameters (the intercept in the purchase quantity model) is 72% and 18.6%, respectively. Because in this case the firm is not targeting on γ_{i4} directly, it is clear that the profit effects of targeting ads to consumers through their ad viewing behavior is generated by the correlation between their ad viewing behavior and their inclination to buy the product. This suggests that consumers with different advertising consumption behaviors provide differential value to firms.

Figure 19 plots the share of ads allocated to a household ($y-$axis) against α_{i4} ($x-$axis) for the benchmark case on the left, and for the ad-viewing allocation case on the right. The probability of watching an ad at baseline for the same households ranges from 0.9 to 0.99. While there is no relationship between ad exposures and ad-viewing behavior in the benchmark case (left plot), this changes considerably after targeting is implemented (right plot). Households are sorted on α_{i4}, and in particular are assigned more exposures when they are more likely to watch the advertisement. The 'S-shaped' curve suggests cutoff rules based on estimates of α_{i4} or the probability of watching the ad, and it may provide useful heuristics for an advertiser seeking to allocate a fixed number of advertisements across consumers.

We now explore how the full information targeting policy ("Targeting: Full Information" case) improves on the ad-viewing policy for brand 4. Figures 20a and b plot the difference in the ad allocation to a consumer under the full information targeting policy relative to the ad-viewing policy ($z-$axis) against

Figure 19: Ad Allocations to Households on Parameter α_{i4}



their $(\alpha_{i4}, \gamma_{i4})$ parameters (Figure a), and against their expected probability of watching the ad and expected demands (Figure b). Looking at Figures 20a and b, we see that the full information policy trades off the probability of watching an ad with the incremental effect of that ad consumption on quantities. In particular, Figure 20a shows there exist some consumers with high α_{i4} (high probability of watching the ad), but with low γ_{i4}; these consumers will not purchase much of the product even though they like the ad. A set of consumers who have low α_{i4} (low probability of watching the ad) and high γ_{i4} also exists. Though these consumers may have a high probability of skipping the ad, conditional on consuming the ad they purchase large quantities. Figure 20a shows the full information policy reallocates ads away from the first type of households towards the second. Looking at Figure 20b, we see that at high expected quantities, ad-allocations are driven primarily by the quantity response, while at low expected quantities, allocations are driven primarily by the ad response. The net effect of these reallocations is a 5.5% (112.1÷106.3) improvement for the full information policy over the ad-viewing only targeting policy for this brand. Overall, we have shown that ad consumption behavior is a valuable metric for allocation of exposures by firms, and that it can capture a significant portion of rents when compared to both the status quo and a full information ad targeting benchmark.

6.3 Results: Welfare

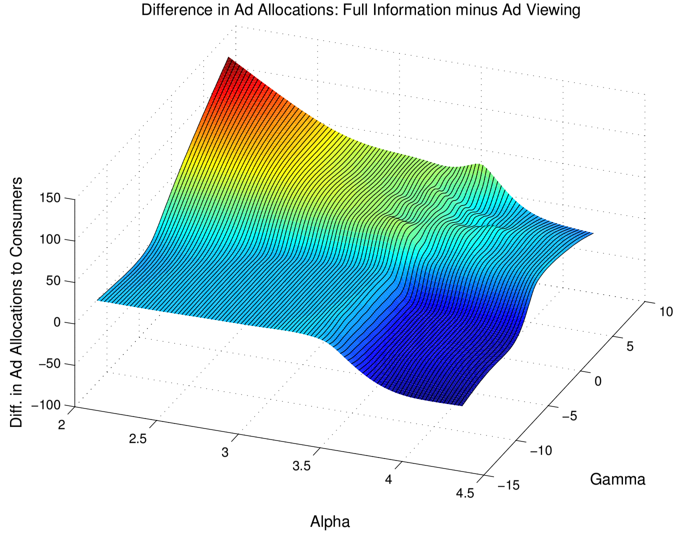
In order to compare the relative gains from targeting on ad-viewing behavior, we consider the total consumer surplus of consumer i, which is given by,

$$Total\,Welfare_i^T = \sum_j \left(\mathcal{E}CS_{ij}^T\right) + E_i \tag{40}$$

where $\mathcal{E}CS_{ijt}^T$ denotes household i's expected consumer surplus from consuming brand j at time t (from equation 25) and E_i is household i's income, which we assume to be fixed during the counterfactual time

Figure 20: Comparing "Full Information" Ad-Targeting to "Ad-Viewing" Targeting

(a) Ads as a function of $(\alpha_{i4}, \gamma_{i4})$



(b) Ads as a function of expected demand and ad-watching

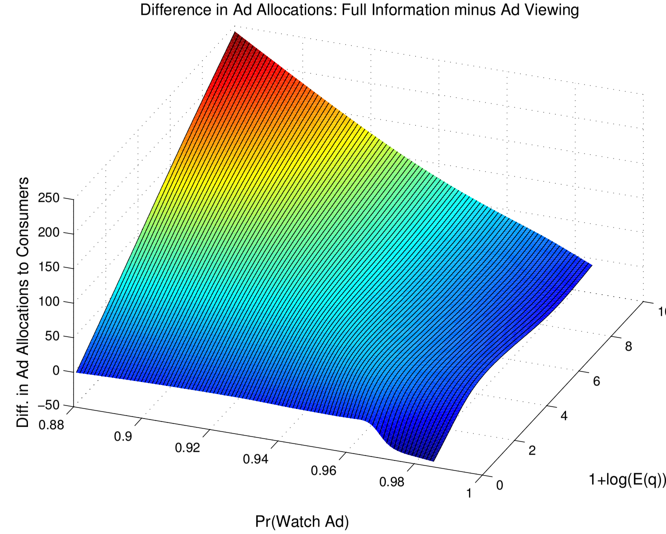


Table 14: Relative Consumer Welfare under Counterfactual Scenarios

Brand:	2	3	4	5	7	8	9	10	11
Targeting holding prices fixed	0.07	0.15	0.55	-1.16	0.04	-0.16	0.09	0.84	0.01
Targeting allowing prices to adjust	-1.24	0.05	0.09	-1.44	-1.46	-0.19	-1.26	-0.92	-0.02

period. Expression (40) captures the total consumption welfare of household i, which is the statistic of interest. Rather than assume income values for the households in the sample, we derive an income-invariant statistic to assess how ad-viewing based targeting performs relative to the full targeting technology. The share of welfare provided by ad-viewing based targeting ($Share\,Welfare$) is given by,

$$Share\,Welfare = \frac{\sum_i \left(Total\,Welfare_i^{Ad\,V.} - Total\,Welfare_i^{Data}\right)}{\left|\sum_i \left(Total\,Welfare_i^{Full} - Total\,Welfare_i^{Data}\right)\right|} = \frac{\sum_i \sum_j \left(\mathcal{E}CS_{ij}^{Ad\,V} - \mathcal{E}CS_{ij}^{Data}\right)}{\left|\sum_i \sum_j \left(\mathcal{E}CS_{ij}^{Full} - \mathcal{E}CS_{ij}^{Data}\right)\right|} \tag{41}$$

where "Data" denotes the ad-allocation in the last quarter in the data, "Ad V." denotes the ad-viewing based ad targeting technology and "Full" denotes the full optimization scenario. The numerator of $Share\,Welfare$ captures the welfare impact of allowing ad-viewing based targeting, over and above the performance of the current targeting practices in the data, and the denominator captures the magnitude of the impact of allowing full ad re-optimization over and above the same benchmark. The numerator and denominator share the same sign, so keeping the denominator positive reveals the direction of welfare.

Table 14 presents the average relative welfare impact by brand. We find that targeting advertisements based on ad viewing behavior generally increases consumer welfare, but can also decrease it.

The underlying reason for the negative impact is that ad exposures interact with the marginal utility of consumption. In this case a consumer who is exposed to an additional ad impression may buy more because of the ad complementarity effects, but derive little additional utility from the incremental products sought. Similarly, a consumer for whom an exposure is canceled may lose more than the first one because even though she now buys only little less, the foregone units provided significant amounts of welfare due to ad complementarities.

The second row of table 14 presents average welfare changes when the firm is allowed to practice linear targeted pricing. Perhaps not surprisingly, in this case consumers always become worse off, although in a few cases they are slightly better off than in the status quo.

The values above reflect the heterogeneity that we find across products in terms of welfare effects. One takeaway is that whether or not more addressable environments will benefit consumers needs to be analyzed on a case by case basis. The reason is that the existence of complementarities introduces complexity in welfare measurement. This complexity arises because by reallocating ad exposures firms affect both the level of demand as well as the utilities of marginal products.

One positive take away from this analysis is that there is potential for consumers to be better off, at least under some counterfactual scenarios. Given that the profit simulations mostly show positive impacts for firms, it therefore seems possible that *both* firms and consumers can be better off in some of the new addressable TV environments.

7 Conclusions

An empirical assessment of the demand for advertising is presented. The model views advertising consumed by an agent as a deliberate choice. Following Becker and Murphy's (1993) theory of complementarities, we assume this choice is co-determined with the choice of consumption of the advertised products. Further, following the theory, joint consumption of advertising and products is allowed to generate complementarities in utility. Using data on detailed product purchase and TV ad-viewing choices by consumers, we document that many ads are skipped, and that the skip-rates are explained by the quantity of the advertised product that was purchase recently. We also document that advertising significantly shifts quantity demanded. These results provide support for a model of complementarities.

We then present and estimate a structural model of demand for products and advertising that allows for such complementarities. Viewing advertising as a deliberate choice by consumers changes the way we assess the "treatment" effect of advertising. Using the model, we document that endogenous advertising effects are important to consider for understanding the way advertising works and for assessing its effects on consumer welfare.

Motivated by the "addressable" future of TV ad-markets in which targeting advertising on the basis of ad-viewing and product purchase behavior is possible, we use the model and estimates to simulate a series of counterfactuals. We simulate how demand, welfare and profits would change if an advertiser could target ads to consumers (a) on the basis of anticipated skipping behavior (which in the presence of complementarities indirectly selects high demand-consumers); (b) on the basis of the full model of ad-and-product demand; and (c) on the basis of the full model of ad-and-product demand while also implementing targeted first-degree price discrimination. We find that profits are higher under all ad and price targeting scenarios considered, but that targeting on the basis of ad-viewing alone makes up for about 16% of total potential increase in profits, suggesting the value of this policy for advertisers. We also find that net consumer welfare can rise in the new targeted environments, primarily derived from the increased surplus accruing to high-volume consumers. These results suggest that it may be possible that firms and consumers are both better off in the new addressable TV environments, though there is considerable heterogeneity across brands on this dimension.

We also believe our results have implications for revenue-models in new ad-driven markets. Monetization of advertising based on the active choices of consumers is increasingly becoming the norm in online markets. For example, YouTube now utilizes an advertising format called TrueView In-Stream in which an advertisement plays first for a few seconds, after which a viewer can choose to skip to the video or watch the rest of the ad. An advertiser using TrueView In-Stream only pays for an impression if the viewer watches a minimum of 30 seconds of the ad before skipping to the intended video (YouTube 2014). Similarly, advertisers pay for Promoted Videos on Twitter only when a user plays the video (Regan 2014). Thus, increasingly, understanding which users choose to consume ads is of relevance to advertisers in digital media. As it becomes easier to track which ads are skipped and which are watched to completion, it may become possible to better understand consumers' preferences for advertisements, and to relate them to preferences for products like we do here.

A limitation of the current analysis is the perfunctory treatment of ad content (except for the brand of the message). Unfortunately, data on the content of the ads could not be obtained from the corporate sponsor in this study to facilitate this analysis. Our data does contain a copy id unique to each creative. In regressions of the percent of ad watched on a set of copy id fixed effects, we found many of the copy id fixed effects significant suggesting that some creatives are watched significantly longer than others. This persists when including household fixed effects (to control for the fact that the kinds of households viewing different copy ids may be different) and brand fixed effects, suggesting that there is variation in propensity to watch within ads for a given brand. We view studying the role of advertising content in ad-skipping and consumption to be an important area of future research. Finally, the ad consumption data used in this study, while novel, contains measurement error. New technology that better captures consumer attention can help make the analysis sharper.

8 References

- Ackerberg, D. (2001). "Empirically Distinguishing Informative and Prestige Effects of Advertising," *RAND Journal of Economics*, 32(2), 316-333.
- Allenby, G.M. and Rossi, P.E. (1999). "Marketing Models of Consumer Heterogeneity," *Journal of Econometrics*, 89, 57-78.
- Anderson, J. R. (1983), "The Architecture of Cognition," Cambridge, MA: Harvard University Press.
- Anderson, T.W and Rubin, H. (1949). "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *The Annals of Mathematical Statistics* 20-1, pg. 46-63.
- Angrist, J. and Krueger, A. (1991). "Does Compulsory School Attendance Affect Schooling and Earnings?" *The Quarterly Journal of Economics*, vol. 106 (4), pg. 979-1014
- Angrist, J. , G.W. Imbens, and A. Krueger. (1999). "Jackknife Instrumental Variables Estimation," *Journal of Applied Econometrics*, 14, 57–67.
- Ashenfelter, O., Krueger, A.B.., (1994). "Estimates of the Economic Returns to Schooling from a New Sample of Twins," *American Economic Review*, 84 (5), 1157–1173.
- Auld, D. A. L. (1974). "Advertising and the Theory of Consumer Choice," *The Quarterly Journal of Economics*, 88(3), pg. 480-87.
- Ausubel, L. (1999). "Adverse Selection in the Credit Card Market," mimeo, University of Maryland.
- Bagwell, K. (2007). "The Economic Analysis of Advertising," *Handbook of industrial organization*, 3, 1701-1844.
- Becker, G. and Murphy, K. (1993). "A Simple Theory of Advertising as a Good or a Bad," *The Quarterly Journal of Economics*, 108(4), 941-964.
- Bertrand M, Karlan DS, Mullainathan S, Shafir E, Zinman J. (2010). "What's advertising content worth? Evidence from a consumer credit marketing field experiment," *Quarterly Journal of Economics*. 125:263–305.
- Bhat, C. (2005). "A Multiple Discrete-Continuous Extreme Value Model: Formulation and Application to Discretionary Time-Use Decisions," *Transportation Research Part B*, 39(8), 679-707.
- Biddle, J. and D. Hamermesh (1990). "Sleep and the Allocation of Time," *Journal of Political Economy*, 98(5), 922—43.
- Bound, J., D. Jaeger, and R. Baker. (1995). "Problems with Instrumental Variable Estimation when the Correlation Between the Instruments and the Endogenous Explanatory Variables is Weak," *Journal of the American Statistical Association*, 90: 443–450
- Bronnenberg, B., Dubé, J-P., Mela, C. (2010). "Do DVRs Influence Consumers' Brand Purchases?" *Journal of Marketing Research*, 47(6), Dec.
- Butters, G. (1977), "Equilibrium Distributions of Sales and Advertising Prices," *Review of Economic Studies,* 44(3), 465–91.
- Card, D., (2001). "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica,* 69 (5), 1127–1160

- Chintagunta, P., Dubé, J-P., and Goh, K. Y. (2005). "Beyond the Endogeneity Bias: The Effect of Unmeasured Brand Characteristics on Household Level Brand Choice Models," *Management Science*, 51(5), 832-849.
- Chintagunta, P. and Nair, H. (2011). "Discrete Choice Models of Consumer Demand in Marketing," *Marketing Science*, 30(6), 977-996.
- DellaVigna, S. and Gentzkow, M. (2010). "Persuasion: Empirical Evidence"*, Annual Review of Economics*, September, Vol. 2, pp. 643-669.
- Deng, Y. and Mela, C. (2015). "TV Viewing and Advertising Targeting," working paper, Fuqua School of Business, Duke University.
- Ehrlich D, Guttman I, Schonback P, Mills J. (1957). "Postdecision exposure to relevant information," *Journal of Abnormal Social Psychology*. 54(1):98–102.
- Fox, J. and Lazatti, N. (2016). "Identification of Discrete Choice Models for Bundles and Binary Games," working paper, Rice Univ., Dept. of Economics.
- Gentzkow, M. (2007). "Valuing New Goods in a Model with Complementarity: Online Newspapers," *The American Economic Review*, 97(3), 713-744.
- Grossman, G. and C. Shapiro (1984), "Informative Advertising with Differentiated Products," *Review of Economic Studies,* 51(1), 63–81.
- Interactive Advertising Bureau. (2016). "2015 Internet Advertising Revenue Full-Year Report," Available at: `http://www.iab.net/research/industry_data_and_landscape/adrevenuereport`
- Imbens, G. and Angrist, J. A. (1994). "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62 (2), pp. 467-475.
- Interian, Y., Dorai-Raj, S., Naverniouk, I., P. J. Opalinski, Kaustuv and Dan Zigmond. (2009a). "Ad Quality On TV: Predicting Television Audience Retention," Proceedings of International workshop on Data Mining and Audience Intelligence for Advertising (ADKDD), Paris. Available at: `http://static.googleusercontent.com/media/research.google.com/en/us/pubs/archive/35368.pdf`
- Interian, Y., K, Naverniouk, I., Opalinski, P. J., Dorai-raj, S. and Zigmond, D. (2009b). "Do Viewers Care? Understanding the Impact of Ad Creatives on TV Viewing Behavior," *Re:Think09*,. Available at: `http://static.googleusercontent.com/media/research.google.com/en/us/ pubs/archive/35167.pdf`
- Isen, A. M. (1992), "The Influence of Positive Affect on Cognitive Organization: Some Implications for the Influence of Advertising on Decisions About Products and Brands," in Advertising Exposure, Memory, and Choice, Andrew A. Mitchell, ed. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Kantar Media (2012). "RapidMeter: A Simple Innovation Bringing Audience Measurement to New Markets". Kantar News and Views. 11 July. Available at: `http://www.kantar media-audiences.com.` (5 September 2014).
- Keller, K. (1983). "Conceptualizing, Measuring and Managing Customer-Based Brand Equity," Journal of Marketing, Vol 57(1), pp. 1-22.
- Kim, K., Allenby, G.M. and Rossi, P.E. (2002). "Modeling Consumer Demand for Variety," *Marketing Science*, 21(3), 229-250.
- Lowenstein, G. and J. Elster (1992). "Utility from Memory and Anticipation," In *Choice over Time*, New York: Russell Sage Foundation, p.g 213-34.
- Lee, S., Kim, J. and Allenby, G.M. (2013). "A Direct Utility Model for Asymmetric Complements," *Marketing Science*, 32(3), 454-470.

- McFadden, D. (1974). “Conditional Logit Analysis of Qualitative Choice Behavior,” in P. Zarembka, ed., *Frontiers in Econometrics*, Academic Press, New York, 105-142.
- Mills J. (1965). “Avoidance of dissonant information,” *Journal of Personal Social Psychology*, 2(4):589–93.
- Naik, P. A., M. K. Mantrala, A. G. Sawyer. (1998). “Planning Media Schedules in the Presence of Dynamic Advertising Quality,” *Marketing Science*, 17(3), 214–235.
- Nelson, P. (1970). “Information and Consumer Behavior,” *Journal of Political Economy*, 78(2), 311-329.
- Nelson, P. (1974). “Advertising as Information,” *Journal of Political Economy*, 82(4), 729-754.
- O’Connor, P. (2014). “Political Ads Take Targeting to the Next Level,” *The Wall St. Journal*, Nov. 20.
- Perlberg, S. (2014), “Targeted Ads? TV Can Do That Now Too”, *The Wall St. Journal*, July 14.
- Prelec, D. and G. Lowenstein. (1998). “The Red and the Black: Mental Accounting of Savings and Debt.” *Marketing Science*. 17, 4-28.
- Peterson, T. and Kantrowitz, A. (2014). “The CMO’s Guide to Programmatic Buying”, May 19, Available at: `http://adage.com/article/digital/cmo-s-guide-programmatic-buying/293257/`
- Regan, D. (2014). “Introducing Promoted Video on Twitter,” Available at: `https://blog.twitter.com/2014/introducing-promoted-video-on-twitter`
- Revelet, D. and Train, K. (2001). “Customer-Specific Taste Parameters and Mixed Logit: Households’ Choice of Electricity Supplier,”working paper, University of California, Berkeley.
- Schweidel, D. A., R. J. Kent. (2010). “Predictors of the Gap between Program and Commercial Audiences: An Investigation Using Live Tuning Data,” *Journal of Marketing,* 74(3) 18–33.
- Simester D, Hu J, Brynjolfsson E, Anderson E. (2009). “Dynamics of Retail Advertising: Evidence from a Field Experiment,” Economic Inquiry, 47(3).
- Shafir, E., Simonson, I., & Tversky, A. (1993), “Reason-based Choice,” *Cognition*, 49 (2), pg. 11-36.
- Siddarth, S., A. Chattopadhyay. (1998). To Zap or Not to Zap: A Study of the Determinants of Channel Switching During Commercials,” *Marketing Science,* 17(2) 124–138.
- Stock, J., Wright J. , and M. Yogo (2002), “A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments,” *Journal of Business and Economic Statistics, 20, 518-529.*
- Tellis GJ, Chandy RK, Thaivanich P. (2000). “Decomposing the effects of direct advertising: Which brand works, when, where, and how long?,” *Journal of Marketing Research*. 37(1):32–46.
- Teixeira, T. S., M. Wedel, R. Pieters. (2010). “Moment-to-moment Optimal Branding in TV Commercials: Preventing Avoidance by Pulsing,” *Marketing Science,* 29(5) 783–804.
- Unitam (2009). “Unitam Technology: UNITAM for Sweden,” Available at: `http://www.mms.se`
- Varian, H. (1992), “Microeconomic Analysis,” W.W.Norton and Co., 3rd Ed.
- Vranika, S. (2012). “Ad-Spending Outlook Dims,” *The Wall Street Journal*, (Dec 2).
- Wales, T. and Woodland A. (1983). “Estimation of Consumer Demand Equations with Binding Non-Negativity Constraints.” *Journal of Econometrics*, 21.

- Wilbur, K. (2008). “A Two-Sided, Empirical Model of Television Advertising and Viewing Markets,” *Marketing Science*, 27 (3): 356-378.
- Wyer, R. S., Jr. and Thomas K. Srull (1989), “Person Memory and Judgment,” Psychological Review, 96 (1), 58- 83.
- YouTube (2014). “Video ads move people to choose you,” Available at: `http://www.youtube.com/yt/advertise/why-it-works.html`
- Zigmond, D., Dorai-Raj, S., Interian, Y., Naverniouk, I. (2009). “Measuring Advertising Quality on Television Deriving Meaningful Metrics from Audience Retention Data,” *Journal of Advertising Research*, 49, 419-428

Appendices: For Online Publication

A Ad Skipping and Observable Demographics

Deng and Mela (2015) report that household demographics are not a strong predictor of the observed variation in ad skipping behavior in their TiVo data. Zigmond et al. (2009) similarly reports that a "user-behavior model" which predicts ad-skipping rates using the observed skipping behavior of viewers an hour before the airing of a show performs better than models that use only network, weekday, day-part, and ad duration variables, or models that use only demographics.

In this appendix, we analyze how ad exposure and ad skipping are correlated with observed household demographics in our data. As shown in Table 15, we find that larger households tend to be exposed to more ads, but all else equal, household size does not correlate with skip rates. Homeowners, people over 50 and those with higher levels of income and education tend to see fewer exposures and have higher skip rates. The fact that wealthier, more educated people are more likely to skip an ad is consistent with the interpretation of the cost of an advertisement as the opportunity cost of one's time. Like Deng and Mela and Zigmond et al., we find that observed household demographics explain little of the variation in ad skipping.

B Price Series Construction

The purchase data records the price paid and package volume of transactions at the barcode level. We only observe the prices of purchased products, but in order to estimate the model, we need to reconstruct the price series of the alternatives that were not purchased. Additionally, our model is at the brand level, so we need to transform barcode level prices into brand level prices. Our approach is to reconstruct a barcode-store-week level price series using all observed transactions and weight by purchase volume to create a brand-store-week price per unit. As we do not model store choice, the final step in the price series construction is to create a household-brand-week level price series by creating a weighted average of the store-brand-week price series using the frequency of a household's store visits as the weights. The steps below describe the process used to construct the price series.

1. Although we only observe purchase and TV advertising data for 6,552 households, we observe purchase data for 22,670 households. The entire purchase database is used in the construction of the price series.

2. We observe at least one purchase of 58 different brands in the transaction data. In order to make the model more tractable, we restrict the analysis to the set of brands that have the largest purchase market shares. We focus on the brands that collectively cover 90% of the market. The "other" or smaller brands category has the largest purchase market share (61.72%). These brands generally do not advertise, and because we cannot be sure whether the ads we do observe in the data correspond to the same brands that were purchases in this category, we do not include the "other" brands in the analysis. This leaves us with 12 brands. The remaining brand with the largest market share is brand 195 with 8.49% of all purchases observed in the database.

3. We keep transactions for barcodes that make up at least 5% of brand sales. Brand 839 does not have any barcodes that have at least 5% of sales, so this brand is dropped from the analysis.

4. We keep transactions for stores that have at least 1,000 purchases of the barcodes identified in step 3.

5. A barcode-store-week level price series is constructed by taking the median/mode observed price of the transactions in that week. If there are not any observed purchases in a week, we fill in the median/mode observed price from the previous week. If no transactions are observed to date, we fill in the median/mode observed price in the week of the first observed transaction.

Table 15: Regression of Household Ad Exposure and Ad Skip Rate on Observed Characteristics

	TV Ad Exposures	HH Skip Rate
Income	-51.2714***	0.0092***
	(7.3563)	(0.0024)
Unemployed	-0.9489	0.0178***
	(14.0883)	(0.0046)
Part Time Employed	-22.0769	-0.0049
	(14.3128)	(0.0047)
Higher Education	-104.980***	0.0289***
	(13.5961)	(0.0044)
Age 29 and Under	-37.7855	0.0013
	(24.1770)	(0.0079)
Age 55 and Over	-49.3056***	0.0290***
	(15.1363)	(0.0049)
Children	-45.1231**	-0.0007
	(19.0153)	(0.0062)
HH Size	85.5733***	-0.0074
	(8.0448)	(0.0026)
Urban	6.4441	0.0054
	(12.7932)	(0.0042)
Homeowner	-85.2041***	0.0123***
	(12.4223)	(0.0040)
Constant	360.777***	0.0875***
	(23.018)	(0.0054)
R-Squared	0.0902	0.0635
Observations	4,221	4,221

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note: Income is a categorical variable taking on values of 1, 2, 3, and 4 for increasing levels of household income. *Unemployed* and *Part Time Employed* are dummy variables indicating employment status. *Higher Education* is a dummy variable indicating some education beyond the high school level. The *Under 29* and *Over 55* dummies indicate the age of the head of household. *Children* is a dummy variable recording whether there are children under the age of 18 living in the home. *Household Size* records the number of people in the household. *Urban* is a dummy variable indicating a town population larger than 100,000. *Homeowner* is a dummy variable indicating the residence is owned.

6. Because barcodes correspond to different package volumes, we create a barcode-store-week level price per unit by dividing weekly prices by package volume.

7. A brand-store-week level price series is constructed by weighting the barcode level price per unit by the total volume of that barcode bought at that store over the sample.

8. Finally, we create a brand-household-week price series by averaging the brand-store-week level price series across stores and weighting by household i's count of purchases at store s over the sample period.

9. Using this store weighting procedure, some households do not have a price series for all 11 brands because sometimes a household has never been to a store where a given brand is sold. For example, household 135,075 only ever purchases at store 1. Brand 195 is not sold at store 1.

   The idea behind weighting the store price series by frequency of store visits is intended to reflect a household's best belief about a brand's current price. In the instances where a household never visits a store where a brand is sold, we assume that that household's beliefs about the price of that brand will be equal to the maximum price for that brand across all stores that week.

C Within Store Price Variation Over Time

Figures 21 and 22 in this appendix documents the extensive price variation over time at the brand-level within stores.

Figure 21: Time Series Plots of Store-Level Prices

Figure 22: Time Series Plots of Store-Level Prices - Contd.

D First Stage IV Regressions

This appendix presents the first stage IV regressions for both the specifications using store level price series and those using household price series as instruments (Tables 16 and 17). In all specifications the coefficient on own-price is negative and highly significant.

Table 16: Regressions of Quantity on Price Instruments − Stores 8 and 9 Price Series

	(1) Quantity		(2) Quantity		(3) Quantity	
	Store 9	Store 8	Store 9	Store 8	Store 9	Store 8
Own Price	-0.256***	-0.036*	-0.255***	-0.036*	-0.252***	-0.042**
	(0.046)	(0.021)	(0.046)	(0.021)	(0.045)	(0.021)
Price_{34}	-0.383***	0.019	-0.380***	0.019	-0.340***	0.027
	(0.103)	(0.062)	(0.103)	(0.062)	(0.101)	(0.060)
Price_{54}	-0.015	0.119*	-0.016	0.117*	-0.030	0.112*
	(0.083)	(0.062)	(0.083)	(0.062)	(0.082)	(0.062)
Price_{195}	0.270***	0.041	0.271***	0.040	0.281***	0.039
	(0.049)	(0.029)	(0.049)	(0.029)	(0.048)	(0.029)
Price_{847}	-0.118***	-0.058***	-0.118***	-0.058***	-0.116***	-0.067***
	(0.027)	(0.022)	(0.027)	(0.022)	(0.026)	(0.022)
Price_{852}	-1.765***	0.351***	-1.769***	0.354***	-1.704***	0.364***
	(0.576)	(0.119)	(0.576)	(0.119)	(0.567)	(0.120)
Price_{856}	0.008	0.021	0.008	0.021	0.012	0.017
	(0.010)	(0.014)	(0.010)	(0.014)	(0.010)	(0.014)
Price_{860}	-0.124**	0.013	-0.122**	0.013	-0.110**	0.003
	(0.049)	(0.020)	(0.049)	(0.020)	(0.048)	(0.020)
Price_{865}	-0.089	0.197**	-0.095	0.199**	-0.080	0.157*
	(0.156)	(0.085)	(0.156)	(0.085)	(0.153)	(0.084)
Price_{866}	-0.040	-0.036	-0.040	-0.035	-0.012	-0.031
	(0.098)	(0.036)	(0.098)	(0.036)	(0.093)	(0.035)
Price_{870}	-0.005	0.097	-0.003	0.098	-0.012	0.106
	(0.044)	(0.164)	(0.044)	(0.164)	(0.044)	(0.161)
Price_{907}	0.103**	0.135***	0.104**	0.135***	0.100**	0.124**
	(0.050)	(0.050)	(0.050)	(0.050)	(0.049)	(0.049)
Observations	1,436,400		1,436,400		1,436,400	
First Stage F	6.10		6.13		5.79	
Week FE	Y		Y		Y	
Brand FE	Y		Y		Y	
Store FE	N		Y		N	
HH FE	N		N		Y	

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Estimates from first stage IV regression of quantity on price instruments. The dependent variable quantity records the cumulative package volume household i purchased of brand j in the 14 days preceding day t. Quantity is instrumented using the price series for all brands at stores 8 and 9, the two most frequently visited stores in the data.

Table 17: Regressions of Quantity on Price Instruments − Household Price Series

	(1) Quantity	(2) Quantity	(3) Quantity
Own Price	-0.238***	-0.235***	-0.236***
	(0.026)	0.026	(0.026)
$Price_{34}$	0.008	0.035	0.033
	(0.055)	(0.057)	(0.051)
$Price_{54}$	-0.142***	-0.149***	-0.159***
	(0.049)	(0.049)	(0.046)
$Price_{195}$	0.058**	0.048*	0.021
	(0.024)	(0.026)	(0.024)
$Price_{847}$	-0.017	-0.027*	-0.031**
	(0.015)	(0.014)	(0.013)
$Price_{852}$	0.221*	0.098	0.037
	(0.116)	(0.097)	(0.080)
$Price_{856}$	0.015**	0.013**	0.015**
	(0.006)	(0.006)	(0.006)
$Price_{860}$	0.062**	0.067***	0.009
	(0.025)	(0.026)	(0.023)
$Price_{865}$	0.039	0.081*	0.134***
	(0.049)	(0.047)	(0.050)
$Price_{866}$	0.016	0.020	-0.024
	(0.040)	(0.038)	(0.036)
$Price_{870}$	-0.007	0.002	0.062
	(0.051)	(0.052)	(0.048)
$Price_{907}$	-0.012	0.001	0.016
	(0.032)	(0.032)	(0.030)
Observations	1,436,400	1,436,400	1,436,400
First Stage F	9.41	9.48	8.84
Week FE	Y	Y	Y
Brand FE	Y	Y	Y
Store FE	N	Y	N
HH FE	N	N	Y

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Estimates from first stage IV regression of quantity on price instruments. The dependent variable quantity records the cumulative package volume household i purchased of brand j in the 14 days preceding day t. Quantity is instrumented using a household-specific price series for each brand. The household-price series is constructed by weighting the store level price series for each brand by the percentage of purchases made by that household at each store over the duration of the data.

E Evidence that Advertising Shifts Quantities Purchased

This appendix reports on several specifications and data cuts to document that advertising has a positive effect on product demand.

First Cut: Cross-sectional Analysis Relating Advertising and Product Demand We start by checking whether households who view more advertisements also purchase more on average. At a minimum, support for a model with complementarities requires seeing a positive covariation between quantities and ads in the data. We explore the joint distribution of total quantity purchase and total category ad consumption at the household level for different levels of ad consumption. To do this, we split the sample of households into three buckets − the lowest quartile, the middle two quartiles, and the upper quartile of the distribution of total ad consumption over the panel duration.[19] These buckets correspond to households who viewed between 0 and 65 ads, between 65 and 448 ads, and 448+ ads, respectively. Then, we estimate the density of purchase quantities for each of these groups. Table 18 summarizes the estimated kernel distributions. The quartiles of the purchase quantity distribution are larger for households in higher ad quartiles. Two-sample Kolmogorov-Smirnov tests reject the null hypotheses that these samples come from the same distribution.[20]Panel Analysis

Based on this cross-sectional analysis, we cannot conclude that advertising *per se* induces the shift out in the purchase density. We can use the panel variation to test if within-household variation over time in purchase quantity for a brand is related to cumulative past advertising consumption by that household of that brand.

We define cumulative past advertising consumption as the sum of the percentage watched of the advertisements to which the consumer was previously exposed. We construct this variable for the preceding 1, 2, 3, and 4 weeks, and regress household i's day t purchase quantity of brand j on household i's cumulative past advertising consumption of ads for brand j. Each observation in the regression is a household-brand-day. This regression is estimated unconditional on purchase, meaning that we include days with no-purchase in the analysis setting quantity equal to 0. We also control for the price per unit of brand j. Because we only observe prices when a purchase is made, we reconstruct the price series for the 11 most frequently purchased brands in the data and restrict all our analyses to these brands. Appendix A describes in detail how we constructed the price series for these brands.

Advertising is endogenous in this regression. Unfortunately, we do not have an instrument that moves ad consumption independently, which can be excluded from the propensity to buy more units. Given the lack of co-ordination of national TV advertising with local determinants of purchases that we described in section 3, we believe the main source of endogeneity concerns is from unobserved heterogeneity and seasonality. The heterogeneity concern reflects the confound that consumers who like a brand also watch more of its ads, and the seasonality concern the confound that consumers may buy more and watch more ads during holiday seasons.

Given we have panel data over a long time horizon, we can include very rich controls for both, that address to a large extent these concerns. We include *household-brand* fixed effects to control for unobserved heterogeneity. Thus, our coefficients are estimated off within household-brand variation over time rather than across household variation and across brand variation. We also include *week* fixed effects in order to address the concern that there may be unobserved, time-varying shocks driving both purchases and ad consumption that remain even after including household-brand fixed effects. In particular, we estimate the following specification,

$$q_{ijt} = \beta_{0ij} + \beta_{0t} + \beta_1 A_{ijt} + \beta_2 p_{ijt} + \epsilon_{ijt} \tag{42}$$

where q_{ijt} is daily purchase quantity in equivalent units and A_{ijt} is a cumulative ad-duration variable (defined more precisely in Table 19). Table 19 presents the results. Consistent with our cross-sectional findings, the effect of cumulative past advertising consumption is positive and statistically significant across all time windows we consider. To interpret the magnitudes of the ad-effect, we also report in the last row the effect on daily quantity demanded of a 1 SD increase in the cumulative ad consumption variables over the past 1, 2, 3, and 4 weeks. Across specifications, we find that a 1 SD increase in ad

[19]For the remainder of the regressions reported here, we restrict our analyses to only include the households who made at least one purchase and were exposed to at least one ad.

[20]We reject the null hypothesis that the observed purchase quantities for households in the bottom quartile and middle two quartiles of the ad consumption distribution are drawn from the same distribution ($p \approx 0$); we also reject the null for the comparison between the middle two quartiles and the upper quartile of the ad consumption distribution ($p \approx 0$).

consumption over the past 1-4 weeks increases the mean daily quantity demanded by 3.4−4.5%. For instance, looking at the last 3 rows of Table 19, the mean daily quantity demanded is 7.93 equivalent units. A 1 SD increase in $A_{ijt,7}$− the ad consumption over the last one week − increases the mean daily quantity demanded by 3.38%.

Even though we have included a rich set of controls, concerns may remain about individual-specific correlated time-varying shocks to both product and ad consumption. One consideration is a form of "activity bias" − that the consumer is busy during some time periods, so is likely both to purchase less and skip ads, which manifests as spurious co-movement in joint consumption. For example, when a consumer goes out of town, we might observe zero purchases and zero ad consumption, which could create spurious correlation between purchase quantity and ad consumption. To assess this, we re-estimate the same model, restricting the data to days in which a household purchased at least one brand and is thus observed to be "active" in the data. We use our preferred specification in which past ad consumption is defined over the preceding two weeks (Column 2, Table 19). Again, we continue to estimate a positive relationship between purchase quantity and cumulative ad consumption (see Table 20).

Table 18: Conditional Distribution of Purchase Quantity by Ad Consumption Quartile

	Purchase Quantity		
Ad Consumption	25th Percentile	50th Percentile	75th Percentile
$0<A\leq 65$	7.78	19.54	41.30
$65<A\leq 448$	11.42	25.90	53.38
$A>448$	13.54	31.54	62.49

Note: Purchase quantities reported in 1,000's of equivalent units.

Table 19: Regression of Daily Purchase Quantity on Cumulative Ad Consumption

	(1) Quantity	(2) Quantity	(3) Quantity	(4) Quantity
$A_{ijt,7}$	0.2119***			
	(0.0337)			
$A_{ijt,14}$		0.1251***		
		(0.0209)		
$A_{ijt,21}$			0.0888***	
			(0.0158)	
$A_{ijt,28}$				0.0920***
				(0.0138)
Price Per Unit	-12.727***	-11.993 ***	-10.670***	-10.920***
	(2.6964)	(2.7439)	(2.7521)	(2.8366)
Observations	29,590,182	29,218,979	28,847,776	28,476,573
HH-Brand FE	Y	Y	Y	Y
Week FE	Y	Y	Y	Y
Ave Quantity	7.93	7.95	7.96	8.02
Effect of +1 SD Ads	0.27	0.28	0.27	0.36
Percent of Mean D.V.	3.38%	3.49%	3.44%	4.49%

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: $A_{ijt,\tau}$ records the cumulative time household i spent watching ads for brand j in the τ days preceding day t. Robust standard errors clustered at the household level.

Table 20: Activity Bias Robustness Check of Regression of Daily Purchase Quantity on Cumulative Ad Consumption

	(1) Quantity
$A_{ijt,14}$	2.6847***
	(0.5092)
Price Per Unit	-339.60***
	(21.700)
Observations	883,736
HH-Brand FE	Y
Week FE	Y
Ave Quantity	262.74
Effect of +1 SD Ads	6.23
Percent of Mean D.V.	2.37%

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: $A_{ijt,14}$ records the cumulative time household i spent watching ads for brand j in the 14 days preceding day t. Robust standard errors clustered at the household level.

F Are Complementarities at the Brand or Category Level?

In this appendix we explore whether the complementarities between ad consumption and product consumption occur at the brand or the category level. In column 2 of Table 21 we regress ad consumption for brand j on the quantity purchased of brand j and the quantity purchased of all other brands $-j$. Column 3 includes total product consumption across all brands as the independent variable. Cross-brand effects are estimated to be negative, though not statistically significant. Table 22 runs the reverse regressions of quantity purchased of brand j on cumulative own and cross advertising of all other brands. After controlling for own effects, the cross-effects are not significant. The last column uses cumulative advertising for all brands as the independent variable; this effect is marginally significant. These results suggest that complementarities between product and ad consumption operate at the brand level as opposed to the category level in these data.

Table 21: Regressions of Ad Consumption on Cumulative Quantity

	(1) Percent Ad Watched Conditional on Exposure	(2) Percent Ad Watched Conditional on Exposure	(3) Percent Ad Watched Conditional on Exposure
$\tilde{Q}_{ijt,14}$	0.0010**	0.0010**	-
	(0.0004)	(0.0004)	-
$\tilde{Q}_{i-jt,14}$	-	-0.0001	-
	-	(0.0001)	-
$\tilde{Q}_{it,14}$	-	-	3.22e-05
	-	-	(0.0002)
Observations	1,436,400	1,436,400	1,436,400
HH FE	Y	Y	Y
Brand FE	Y	Y	Y
Week FE	Y	Y	Y
Marginal Effect of an Additional Purchase of Average Size on Expected Percentage Watched			
Ave Percent Ad	97.06%	97.06%	97.06%
Effect of +1 Purchase j	0.10%	0.10%	-
Effect of +1 Purchase -j	-	-0.01%	-
Effect of +1 Purchase	-	-	0.00%
Standard errors in parentheses			
*** p<0.01, ** p<0.05, * p<0.1			

Note: Regressions estimated at the household-brand-day-exposure level. The dependent variable Percent Ad Watched records the percentage of the exposure that was watched and ranges between 0 and 1. Regressions estimated conditional on an exposure. $Q_{ijt,14}$ records the cumulative package volume household i purchased of brand j in the 14 days preceding day t. $Q_{i-jt,14}$ records the cumulative package volume household i purchased of all brands besides brand j in the 14 days preceding day t. $Q_{it,14}$ records the cumulative package volume household i purchased of all brands in the 14 days preceding day t. Robust standard errors clustered at the household level.

Table 22: Regression of Daily Purchase Quantity on Cumulative Ad Consumption

	(1) Quantity	(2) Quantity	(3) Quantity
$A_{ijt,14}$	0.1251***	0.1300***	-
	(0.0209)	(0.0207)	-
$A_{i-jt,14}$	-	-0.0050	-
	-	(0.0049)	-
$A_{it,14}$	-	-	0.008*
	-	-	(0.0047)
Price Per Unit	-11.993 ***	-11.996***	-11.936***
	(2.7439)	(2.7440)	(2.7430)
Observations	29,218,979	29,218,979	29,218,979
HH-Brand FE	Y	Y	Y
Week FE	Y	Y	Y
Ave Quantity	7.95	7.95	7.95
Effect of +1 SD Ads j	0.28	0.29	-
Percent of Mean D.V.	3.49%	3.63%	-
Effect of +1 SD Ads -j	-	-0.06	-
Percent of Mean D.V.	-	-0.74%	-
Effect of +1 SD Ads	-	-	0.11
Percent of Mean D.V.	-	-	1.35%

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: $A_{ijt,14}$ records the cumulative time household i spent watching ads for brand j in the 14 days preceding day t. $A_{i-jt,14}$ records the cumulative time household i spent watching ads for all brands other than j in the 14 days preceding day t. $A_{it,14}$ records the cumulative time household i spent watching ads for all brands in the 14 days preceding day t. Robust standard errors clustered at the household level.

G Ad Consumption Model Sensitivity Estimates

In our model, we consider an ad to be skipped if it is not watched to completion. In this section we explore the sensitivity of our results to different definitions of skipping. Table 23 reports the results of a logit model in which we regress the binary decision of whether to watch an ad on cumulative purchase quantity in the previous two weeks. We consider alternative definitions of ad consumption where an ad is considered skipped if a) less than 100% of the ad is watched (the ad is not watched to completion), b) less than 95% of the ad is watched, and 3) less than 75% of the ad is watched. The regression is estimated at the household-brand-day level and household-brand random effects are included to control for heterogeneity across households. The magnitudes of the coefficients on product quantity are similar, showing that our results are not sensitive to our specific definition of ad skipping.

Table 23: Logit Regression of Ad Watched Dummy on Cumulative Purchase Quantity

	(1) Ad Watched 100%	(2) Ad Watched 95%	(3) Ad Watched 75%
$\tilde{Q}_{ijt,14}$	0.045***	0.041***	0.053***
	(0.013)	(0.014)	(0.016)
Observations	1,436,400	1,436,400	1,436,400
Week FE	Y	Y	Y
HH-Brand RE	Y	Y	Y

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Logit model estimated at the household-brand-day-exposure level, conditional on an ad exposure. The dependent variable is a dummy variable with value 1 if the ad is watched and 0 if the ad is skipped. In column 1, we consider an ad to be watched if 100% of the exposure is displayed. In columns 2, we consider an ad to be watched if at least 95% of the exposure is displayed, and in column 3 if at least 75% of the exposure is displayed. $\tilde{Q}_{ijt,14}$ records the cumulative package volume household i purchased of brand j in the 14 days preceding day t, normalized by the average daily purchase volume, $\bar{q}_t$ = 2,015 equivalent units..

H Additional Ad-Model Parameter Estimates

Tables 24 and 25 below report on estimates for the zero inflated Poisson model for the number of ad exposures observed across households in equation (33), as well as the multinomial model in equation (34) for the brand of each ad exposure.

Table 24: Estimates for Zero-Inflated Poisson Model for the Number of Ad Exposures

π	λ
0.6565	7.9383
(0.0007)	(0.0137)

Table 25: Estimates for Multinomial Model of Ad Exposure Brand Content

Brand	1	2	3	4	5	6	7	8	9	10	11
$\hat{\varphi}_j$	-	0.0961	0.1063	0.1232	0.0741	-	0.1552	0.1496	0.1106	0.0956	0.0892
SE($\hat{\varphi}_j$)	-	(-0.0466)	(-0.0503)	(-0.0564)	(-0.0384)	-	(-0.0679)	(-0.0659)	(-0.0519)	(-0.0464)	(-0.0002)

I Simulation Procedure

Here, we discuss in more detail how we implement the simulations in section 5.1 to measure "long-run" elasticities in the model. For each household we simulate purchase and advertising consumption outcomes over the last 3 months in the data at the observed levels of prices and advertising exposures. We compare the model-predicted results to the results of a second simulation in which we allocate additional exposures to each household. The steps below outline the simulation procedure.

1. Restrict the sample to the set of households who viewed at least one ad for brand j.

2. Allocate the additional ad exposures for brand j to the first household, spreading the additional exposures evenly across days in which the household viewed an ad for brand j.

3. Conditional on the initial observed stock of product consumption, take error draws for the ad-skip model and predict ad consumption decisions for each of the s' exposures for all brands on day t.

4. If an ad is skipped, draw the percentage of the ad watched independently from the observed distribution of ad durations in the data. If an ad is not skipped, set the percentage equal to 1.

5. Update the advertising stock $\overrightarrow{A}$ using the simulated ad percentages in t.

6. Conditional on the observed prices in the data and the advertising stock $\overrightarrow{A}$, take error draws for the product purchase model and predict product consumption for all brands on day t.

7. Update the consumption stock $\overrightarrow{Q}$ using the simulated purchase quantities in t. Set $t = t+1$. Return to step 3.

8. Repeat the forward simulation procedure in steps 3 - 7 for $R = 100$ paths of error shocks and average the statistics of interest (total purchased quantity, ad consumption, and consumer surplus) over all simulations. Repeat this procedure for all households.

J Approximating the Solution to the Pricing Problem

This section discusses how we approximate the solution to the optimal pricing problem in our counterfactuals. When the firm is capable of price discriminating across consumers, it chooses p^*_{ijt} by solving the problem,

$$\max_{p_{ijt\geq 0}} \left(p_{ijt} - c_{jt}\right) \times \mathbb{E}\left[x^*_{ijt}\left(p_{ijt}, A_{jt-1}\right)\right] - cb_{ijt} \tag{43}$$

here $\mathbb{E}\left[x^*_{ijt}\left(p_{ijt}, A_{jt-1}\right)\right]$ is the expected demand from consumer i as defined in equation (21). Differentiating the expression above using Leibniz's integral rule yields the first-order condition with respect to prices for profit maximization,

$$\frac{c_{jt}}{p_{ijt}} + \frac{\exp(2\gamma_0)p_{ijt}(p_{ijt}-c_{jt})}{\exp(2\gamma_j)(1+A_{jt-1})^{2\beta}+\exp(2\gamma_0)p^2_{ijt}} + \frac{c_{jt}\left(\pi - 2\tan^{-1}\left(p_{ijt}(1+A_{jt-1})^{-\beta}\exp(\gamma_0-\gamma_j)\right)\right)}{2(1+A_{jt-1})^{-\beta}\exp(\gamma_0-\gamma_j)p^2_{ijt}} - 1 = 0 \tag{44}$$

We choose c_{jt} to fix the variable margin of the firm at 0.7, i.e., $c_{jt} = 0.7p_{ijt}$. To reduce execution time, we use a first-order Taylor series approximation of the left-hand side of first-order condition (44) in order to solve for the optimal price p^*_{ijt}. In particular we solve the problem,

$$f_{\Omega_{jt}}\left(p_0\right) + f^{'}_{\Omega_{jt}}\left(p_0\right)\left(p^*_{jt} - p_0\right) = 0 \tag{45}$$

for p^*_{jt}, where $f_{\Omega_{jt}}\left(\cdot\right)$ is the left-hand side of the first-order condition (44) given $\Omega_{jt} = \{c_{jt}, A_{jt-1}, \gamma_0, \gamma_j, \beta_j\}$ and p_0 is set to the average price p_{jt} observed in the data for product j on date t.

K Steps for Allocation of Ads to Consumers in Counterfactual Analysis

K.1 Summary

This procedure reallocates advertisements by households efficiently. Because ad exposures cannot be allocated in a fractional manner, we optimize the problem of the firm by assigning blocks of ads among households. For example, if at time t the firm is observed to allocate $1,000$ exposures to 300 households in the data, then the average exposure intensity is equal to $3.\overline{3}$. We solve the targeting problem of the firm allocating an integer number of ads across consumers, i.e., in the case above the firm seeks to allocate 3 - *floor*$(3.\overline{3})$ - advertisements to 333 households (as many households as possible given the ad intensity), and one last ad exposure is assigned to household number 334. The problem of the firm is then to identify, based on advertising consumption and product purchasing patterns, which households to target with advertisements.

K.2 General Steps

Below we outline the basic steps in the procedure. A pseudocode listing follows.

1. At t, calculate initial advertisement and quantity stocks, $A0$ and $Q0$.
2. Define *n_exp* as the total number of advertising exposures at time t, and *n_hh* as the number of households exposed to ads in the data at time t. In order to reallocate (whole) ad exposures, define *mean_exp* as "floor(*n_exp* / *n_hh*)", i.e., the result of the integer division of total exposures by households exposed.
3. Re-define *n_hh* as the number of households that can be assigned *n_exp* ads, given that the first *n_hh*-1 households will be exposed to *mean_exp* ads, and the *n_hh*th household will be assigned *last_ads*. (See pseudocode for definitions.)
4. For each household, calculate the expected daily consumption if it were exposed to *mean_exp* advertisements.
5. Then, for each household, calculate the expected daily consumption if it were not exposed to any advertisements.
6. Calculate the marginal effect of average ad exposure on consumption by differencing the results from steps 4 and 5, and rank households in descending order of ad effects.
7. Starting from the top of the list, assign *mean_exp* ads to the first *n_hh* households. Then, assign *last_ads* to the *n_hh*th household.
8. Calculate expected sales given the assignment, and the resulting expected profits.
9. Update the advertising and quantity stocks $A0$ and $Q0$, but use shock realizations rather than expectations. Set $t = t + 1$ and go back to step 1. Stop at $t = 106$.

With endogenous prices, steps 3 and 4 incorporate total expenditure by consumers rather than just consumption. The firm optimizes over price individually, taking into account each consumer's expected demand. The remaining steps are not affected. Finally, while the firm maximizes *expected* profits, in reality advertising and good consumption take place at particular draws of the random variables. This is especially important for step 9, which ensures that the program is simulated forward by taking realizations of consumption of the ad and of the good, rather than their expectations. Hence, in order to extract expected profits appropriately several paths with different draws of ad and consumption shocks are simulated, and the statistics of interest (consumer surplus and profits) are averaged over those simulations.

K.3 Pseudocode

Variable Definitions:

1. A0 is a vector ($N \times 1$) with each household's initial stock of ad consumption
2. Q0 is a vector ($N \times 1$) with each household's stock of good consumption
3. a_t is a matrix ($N \times T$) with each household's ad exposures in the data

4. q_t is a matrix ($N \times T$) with each household's ad exposures in the data

5. hh_id is a vector ($N \times 1$) with a unique id for each household

6. profit is a scalar initiated at zero

```
// main function

   // calculate ad stocks and n.people exposed
   n_exp = sum(a_t(..,t));
   i_hh = a_t(..,t) > 0;
   n_hh = sum(i_hh);

   // calculate number of people to expose ads to, and n.exposures
   mean_exp = floor(n_exp / n_hh);
   ads_left = n_exp - mean_exp * n_hh;
   n_hh = n_hh + floor(ads_left / mean_exp) + 1;
   last_exp = n_exp - (n_hh - 1) * mean_exp;

   // calculate exp.  marginal effects of showing ads
   // (hh_id is a vector of unique id's for individuals)
   qt_pred0 = calc_Eq(hh_id, A0);
   qt_pred1 = calc_exp_q(hh_id, mean_exp);
   delta_qt = qt_pred1 - qt_pred0;

   // find households with highest expected increment
   mat1 = [delta_qt, hh_id];
   mat1 = sort(mat1, 1); // sort by column 1
   target_hh = mat1(.., 2); target_hh = target_hh(1..n_hh);
   last_target_id = target_hh(n_hh);

   // update advertising and quantity with realizations
   a_t(.., t) = 0; a_t(target_hh, t) = ad_viewing(target_hh, mean_exp);
   a_t(last_target_id,t) = ad_viewing(last_target_id, last_exp);
   A0 = A0 + a_t(.., t) - a_t(..,t-14);
   q_t(..,t) = calc_q(hh_id, A0);
   Q0 = Q0 + q_t(.., t) - q_t(..,t-14);


   // calculate statistics of interest (here profits are exemplified)
   profits = profits + calc_profits();

   // go to next period
   t = t + 1;
// Other Functions:

// function calc_profits()
// updates profit

function calc_profits(){
   // get dollar value
   margin = 0.3 * p_j(t);
   return margin * sum(q_t(..,t));
```

```
}
// function calc_exp_q(hh_id, ad_exp)
// calculates expected q for households with id's
// in hh_id, exposed to ad_exp ads

function calc_exp_q(hh_id, ad_exp){
   Eq_pred = 0;
   for i=1..nsimul{
      n_ads = ad_viewing(hh_id, ad_exp);
      Eq_pred(i) = Eq_pred(i) + calc_Eq(hh_id, A0(hh_id) + n_ads);
   }
   return Eq_pred / nsimul;
}

// function ad_viewing(target_hh, ad_exp)
// predicts a draw of number of ads watched given exposure to ad_exp ads
// Note:  (unif_rand return a vector of uniform r.v.'s with
// the same number of elements as target_hh)

function ad_viewing(target_hh, ad_exp){
   for j=1..ad_exp{
      n_ads = n_ads + (unif_rand(0,1) < pr_watch_ad(target_hh));
   }
   return n_ads;
}

// External Functions:

// function pr_watch_ad();
// calculates the expected ad consumption for each household

// function calc_q(hh_id, Ad_Stock);
// calculates a draw of product consumption for each household

// function calc_Eq(hh_id, Ad_Stock);
// returns E[x*_ijt(p_ijt, A_jt-1)] for individuals in hh_id, according to equation (21)
// in the paper.
```