Complementarities in Consumption and the Consumer Demand for Advertising*

Anna Tuchman
Doctoral Student in Marketing
Stanford GSB

Harikesh S. Nair
Prof. of Marketing
Stanford GSB

Pedro Gardete
Asst. Prof. of Marketing
Stanford GSB

First Draft: Oct 2014; This version: November 27, 2014

Abstract

The standard paradigm in the empirical literature is to treat consumers as passive recipients of advertising by firms, with the level of ad exposure determined by firms' targeting technology and the intensity of advertising supplied in the market. This paradigm ignores the fact that consumers may actively choose their consumption of advertising, thus effectively rendering the level of ad exposure an endogenous choice. Endogenous consumption of advertising is common. Consumers can easily choose to change channels to avoid TV ads, click away from paid online video ads (e.g., TrueView ads on YouTube), or discard direct mail without reading advertised details. Becker and Murphy (1993) recognized this aspect of demand for advertising and argued that advertising should be treated as a good in consumers' utility functions, thereby effectively creating a role for consumer choice over advertising consumption. They argued that in many cases demand for advertising and demand for products may be linked by complementarities in joint consumption. We leverage access to an unusually rich dataset that links the TV ad consumption behavior of a panel of consumers with their product choice behavior over a long time horizon to measure the co-determination of demand for products and ads. The data suggests an active role for consumer choice of ads, and for complementarities in joint demand. To interpret the patterns in the data, we fit a structural model for both products and advertising consumption that allows for such complementarities. We explain how complementarities are identified. Interpreting the data through the lens of the model enables a precise characterization of the sub-population that consumes advertising conditional on exposure, and thereby facilitates a more precise assessment of the treatment effect of advertising under such endogenous non-compliance, and assessments of the value of targeting advertising. To illustrate the value of the model, we compare advertising, price responses and consumer welfare under the model to a setup in which ad consumption is not treated as a choice under the control of consumers. We believe the results have important implications for interpreting ad-effects in empirical work generally, and for the assessment of ad-effectiveness in many market settings.

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

*We thank Kyle Bagwell, Lanier Benkard, Sanjog Misra, Martin Pietz, Rahul Telang, Song Yao, David Zvilichovsky; seminar participants at Stanford University and UCLA-Anderson; participants at the 2014 QME (USC), 2014 Marketing Science (Atlanta), 2014 TADC (LBS), and Economics of ICT (Mannheim) conferences; and especially Peter Rossi for useful comments. We thank the Wharton Customer Analytics Initiative and an anonymous sponsoring firm for generously sharing the data and making it available for academic research. The usual disclaimer applies. Please contact Tuchman (tuchman@stanford.edu), Nair (harikesh.nair@stanford.edu) or Gardete (gardete@stanford.edu) for correspondence.

"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 2013 totaled $175 billion, with Internet advertising totaling $43 billion and TV advertising (Broadcast and Cable TV combined) totaling $74.5 billion in revenues (IAB 2014). 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 central 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 papers 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, 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 see an ad. The framework developed by Becker and Murphy has a lot of intuitive appeal for handling advertising choice for this reason. The Becker-Murphy approach is to treat advertising as an explicit good in consumers' utility functions, thereby effectively 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 advertised product

[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), or the so-

increases the marginal utility from ad consumption, so observed purchase quantities are informative in explaining advertising consumption. Thus, utilizing the theory helps us 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. As we explain in the next section, firstly, a formal assessment of who would consume ads and the extent to which the ad-consumption shifts purchase behavior is important for the individual-level targeting of advertising. Such targeting is increasingly 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 are increasingly allowing advertisers to target advertisements to specific consumers based on their observed historical product purchase and ad-viewing behavior (see Perlman 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 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-s 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 the consumption of targeted advertising and advertised products is key to the targeting problem, and for valuing inventory in such markets. Viewing targeting through the lens of a model with complementarities also changes typical intuition about ad-targeting. The conventional wisdom is that those that 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., $\frac{\partial U}{\partial A} < 0$, $\frac{\partial^2 U}{\partial A \partial Q} > 0$). Becker-Murphy give the example of fashion- and fitness-related ads that makes some consumers feel worse about themselves due to unflattering peer comparisons, but still cause them to buy more cosmetics or fitness-related equipment. The assessment of targeting thus depends on the measurement of complementariness encapsulated in the cross-partials of utility.

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 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.

Second, modeling advertising as a choice changes the assessment of the welfare effects of ads. In our framework, advertising 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 simply avoid them upon exposure. Thus, in our set-up, ad exposures and ad consumption are separate constructs. The typical treatment of advertising assumes ad consumption is the same as ad exposure (because the choice to consume advertising is not modeled). The typical treatment thus induces measurement error into the ad consumption variable that is used to assess the effect of advertising on demand. The measurement error is non-classical because the measurement error in the "X" variable − ad exposures − is correlated with the true "X" (ad consumption), as well as the unobservables driving the "Y" variable (quantities purchased), and the induced bias can go in either direction. When the effect of ads on quantities is biased upward, the standard treatment overstates the welfare effects of ads, and when the effect of advertising is biased downward, the standard treatment understates the welfare effects of ads.

Thirdly, as we explain in more detail in the section below, the fact that advertising is actively chosen by consumers complicates the assessment of the causal effects of advertising by inducing the problem of non-compliance. 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. A formal model of who takes up treatment is then very useful to assess the full distribution of treatment effects, as well as to precisely characterize the sub-populations to which the measured treatment effects apply.

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 of this sort 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 AC Nielsen, a large market-research company. The purchase data records the product bought, day of purchase, its price, package size, number of units purchased, brand, and manufacturer information for each household. The TV advertising data is 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 over a long period of time for a given household. There are over 100,000 purchase occasions and about 1.5M advertising exposures captured in total. No channel, show or network characteristics associated with the ads are available. But, importantly, the data 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.[2] This variable equals one if

[2]These data are collected using Nielsen PeopleMeters. The following excerpt from `http://en.wikipedia.org/wiki/People_meter` (accessed Sept 18, 2014) describes the technology: "A people meter is an audience measurement tool used to measure the viewing habits of TV and cable audiences. The People Meter is a 'box' about the size of a paperback book. The box is hooked up to each television set and is ac-

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 very powerful 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 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 makes the data unique compared to traditional "single-source" advertising panels.

We first use the data to test for evidence that advertising is complementary to consumption. An important implication of the model 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 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 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 ad consumption increases quantities purchased, and that higher quantities purchased increases ad consumption on the margin. 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-choices. The econometric model allows for complementarities, but does not impose them. To identify the model, we leverage the rich variation observed in 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). 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

companied by a remote control unit. Each family member in a sample household is assigned a personal 'viewing button'. It identifies each household member's age and sex. If the TV is turned on and the viewer doesn't identify themselves, the meter flashes to remind them. Additional buttons on the People Meter enable guests to participate in the sample by recording their age, sex and viewing status into the system. For an overview, please see: `http://www.nielsen.com/content/corporate/us/en/solutions/measurement/television.html`.

changes in advertising and product consumption in response to changes in the number of exposures. We find significantly 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. We then simulate the induced change in advertising and product consumption in response to a change in the quality of advertising creative. Again, we find large changes in take-ups and long run quantities purchased. A framework without a model for ad-choice will find it difficult to capture these effects because of a lack of a way to predict how changes in ad-quality affect ad-take up and viewership. Finally, we estimate a benchmark comparison model in which we do not allow for ad-skipping, effectively treating ad consumption as equivalent to exposures. We find that this alternative model overstates advertising effects, and overstates the effect of advertising on quantities. Comparing implied consumer welfare, we find that relative to the full model, consumer welfare is overstated. We are currently working on other counterfactuals in which we assess the gain from micro-targeting by re-optimizing individual-level advertising and pricing based on the model.

We believe our results have implications for how researchers view advertising effects and welfare as discussed above, and also for revenue-models in new ad-driven markets. Monetization of advertising based on the active choices by consumers to view the advertising targeted to them 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 their 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, so as to develop a richer understanding of the demand for advertising and products.

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 through 7 present the estimation and simulation results. Finally, Section 8 concludes.

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

Our approach to measuring advertising effects is to develop and estimate a structural simultaneous equations model of the decision to consume products and advertising, and to 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 fundamental 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. Though $\tilde{d}_i \perp \{y_{i0}, y_{i1}\}$, now d_i is no longer independent of $\{y_{i0}, y_{i1}\}$. The decomposition in (2) no longer obtains and "treatment" is no longer randomized because of non-compliance. 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.

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.

Why would an advertiser or a TV network care about non-compliance? Randomization does after all, 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 who the response from the ad is the 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 facing 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 high-spending TV advertisers in 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]

[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 solicitations offering superior terms.

From the TV network's perspective, non-skippable ads by advertisers may be favored all else equal, because it reduces the chance that consumers may switch away from TV during commercials. Hence, a TV ad-network may be willing to entertain price discounts on ad 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 at 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 is insufficient.

Furthermore, in many advertising contexts, we would also like to know the entire distribution of advertising effects, not just the mean effect of ad exposure as in Equation (2). Due to the linearity of expectations, only the *mean* difference is identified this way (not the median or other quantiles). Distributional effects require additional assumptions. Knowledge that the ad-effect is a particular value for the average unit has no implication for what it is for some sub-population of consumers. In practice, we may be interested in several questions related to those sub-populations. Some of these include: what is the effect of ads for those sub-populations? Are the sub-populations that respond targetable, addressable and costly to acquire (for advertiser)? These issues apply irrespective of whether or not advertising consumption is chosen by the consumer, and are difficult to address without a well-posed model with heterogeneity. A separate question is why do some sub-populations respond and others not (i.e., mechanisms)? 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.

Finally, in the situation where advertising consumption is an explicit choice, a separate issue arises from characterizing the sub-population to which the measured effect applies. In the above example, 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. So in the absence of additional assumptions or information, we cannot say for which groups of consumers the measured effects apply. A model for choice of ad consumption becomes very attractive in this situation because it articulates why and which consumers will choose to consume the ad and thus helps characterize these sub-populations precisely.

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

By placing advertising consumption in the same footing as product consumption, Becker and Murphy's framework provides an elegant framework to handle ad and product choice with a clear link to micro-foundations. Becker-Murphy's theoretical framework has to be modified in four ways when confronting household-level panel data on TV ad consumption as we do in our empirical application. First, it needs to be augmented to allow for stock effects of advertising (as opposed to purely flow effects) when considering the panel level variation over time. Stock effects are required to handle the carryover effects of advertising that have been extensively 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, a common budget constraint over product and advertisement consumption of the type envisaged by Becker-Murphy may not hold in TV ad consumption decision contexts where the product purchase and ad consumption decisions are separated in time.

3 Data Description

As mentioned in the introduction, the dataset used in our empirical analysis comprises a long panel of household-level matched purchase and advertising data from a large sample of households in a Western European country. The data is collected by AC Nielsen. The data covers purchases and advertising exposure and consumption for all brands sold in a product category. The product category is described as a fast moving consumer packaged good that is primarily sold in brick and mortar stores. For privacy reasons the identities of the product category and origin country are not revealed. The sample is not entirely representative in that it slightly over-samples households with elderly people and households that have internet connections. Purchases are recorded at the household-brand-day level and ads are captured at the household-brand-exposure level.[4] Finally, demographic information about the households, including the number of family members, number of children, and average income and

[4]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.

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 1: Summary Statistics

	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.

Table 2: Across Household Variation in Purchases

education levels of the head of household is also available to track heterogeneity.

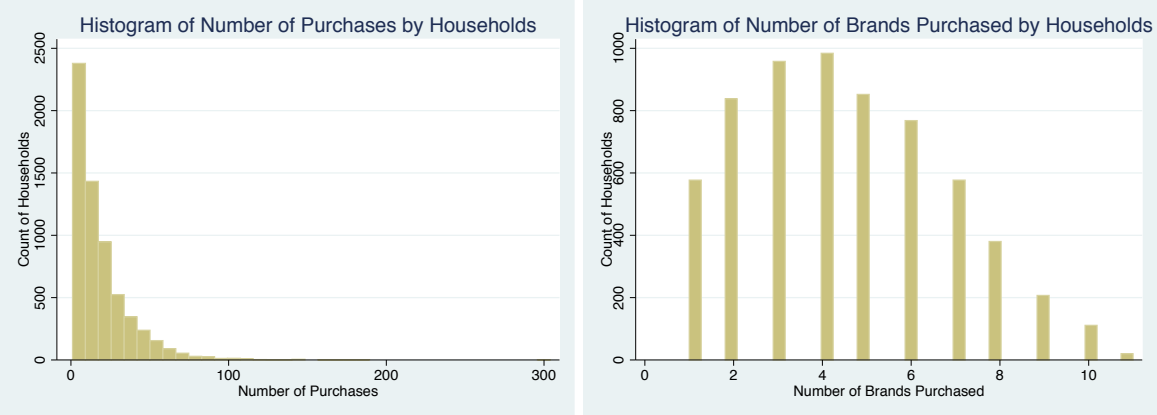


Figure 1: Distribution of Purchases and Number of Brands Purchased by Household

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 full distribution of purchases across households, and Figure 1b shows the variation in the number of brands purchased by different households over the course of the panel. There is significant heterogeneity in both brand preferences as well as the total frequency of purchase.

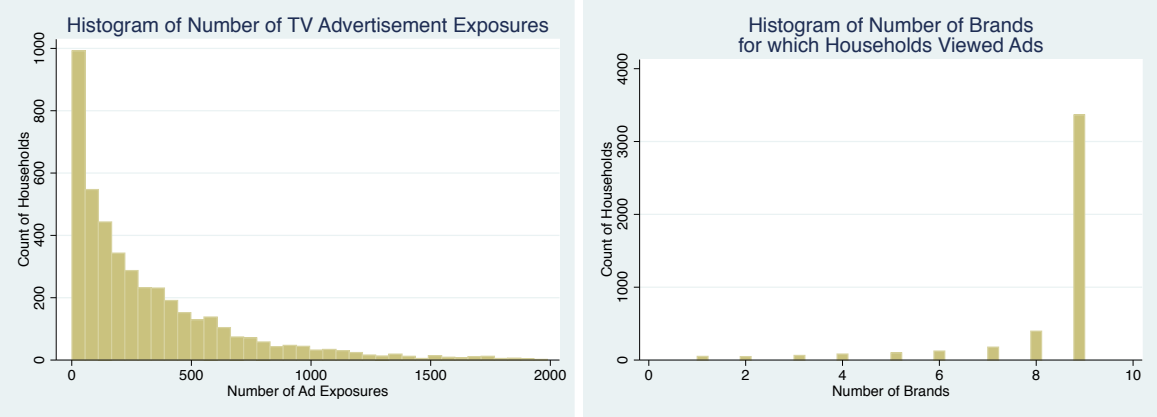


Figure 2: Distribution of Ad Exposures by Household

	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 3: Across Household Variation in TV Advertisement Exposures

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.

Part of the identification of complementarities derives from the extent to which quantity purchased responds to advertising consumption, so it is also interesting to document the variation in quantity purchased in the data. 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). Table 4 reports summary statistics for the quantity purchased on a given purchase occasion defined 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 in day t expressed in equivalent units). Figure 3 shows a histogram of the same variable across households. The mean equivalent volume purchased of a brand is about 2,000 units, and there is extensive variation in purchase quantity across purchase occasions.

Turning to advertising consumption, Figure 4a 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

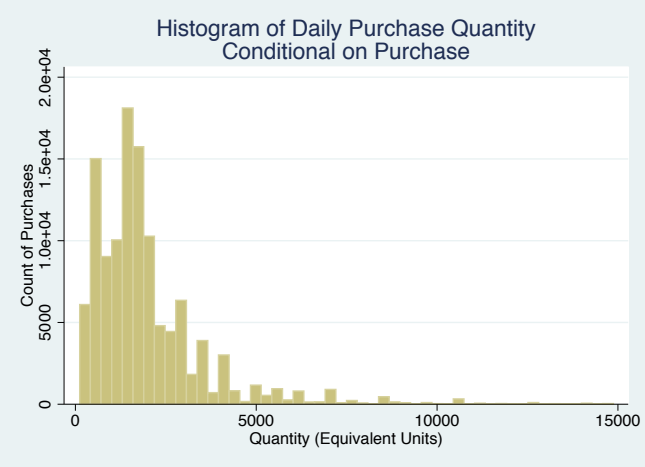


Figure 3: Histogram of Daily Purchase Quantity, Conditional on Purchase

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

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

exposures over the observed length of the panel that are 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 4b 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

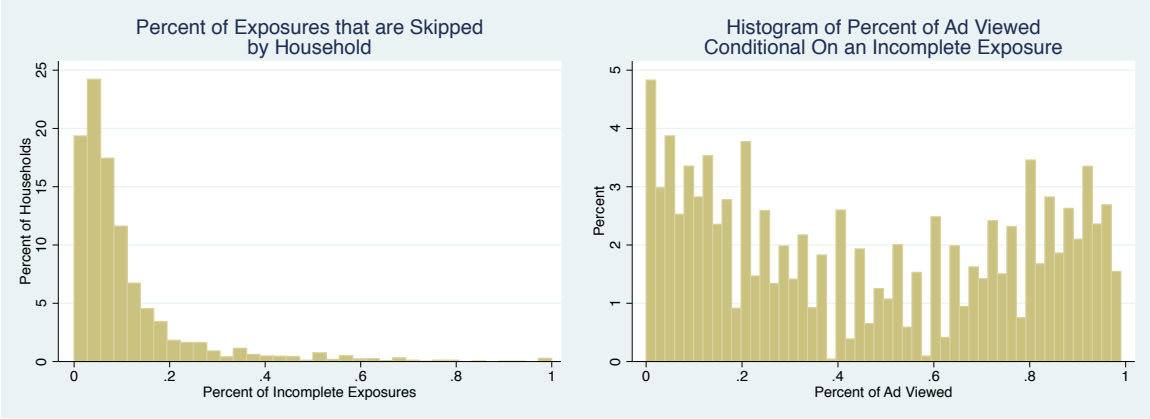


Figure 4: Distribution of Ad Skip Rates and Percentage Watched

variation in how much of an ad is viewed conditional on the decision to skip it.

The 5% ad-skip rate warrants some discussion as it may seem small compared to casual intuition about ad-skipping, especially in relation to online ads.[5] The skip-rates we observe in our TV data are consistent with those reported in the few academic papers we know that have access to data on household-level TV ad-skipping rates, as well as some of the trade press. For instance, Story (2006) reports that prime time shows on Broadcast TV in the US lose roughly about 5% of viewers during commercials. This article quotes a 2005 study by the American Association of Advertising Agencies and the Association of National Advertisers on non-DVR households that found a similar average — 5.6% — of viewers 18 to 49 years old actually skip commercials. Other trade press articles that rely on self-reported survey data often suggest much higher skipping rates (e.g., around 47% in a survey conducted by Jupiter Media reported in Green, 2007), especially for recorded shows played on DVR-s. However, recent academic research that has explored actual TiVo log-data shows that actual ad-skipping rates are much lower than suggested in such self-reported data. Analyzing 46,620 total ad exposures amongst TiVo owning households, Bronnenberg et al. (2010) report that only 3,034 are fast-forwarded (implying a mean skip rate of 6.5%). 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 Zigmod et al. (2009) based on data acquired by Google from the DISH Network in the US, describing 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 that tune away from it without watching it completely − of 1%-3%. These data are more credible than the self-reports used in trade press surveys because they reflect actual ad-skipping behaviors collected in an unobtrusive manner. Our data are consistent with these numbers.[6]

Finally, in more recent research using data from TiVo logs, Deng (2014, Table 1) analyzes the extent to which factors such as 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 explain the variation in ad-skipping across exposures. She finds 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

[5] 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.

[6] A related question is the extent to which observed skip rates are understated by measurement error. For instance, it could be that households lose attention or simply look away when an ad is playing, a form of non-consumption that is not captured by Nielsen's Peoplemeters. This kind of measurement error is possible and our results should be seen with this caveat, though it is likely to be of second-order as it takes the form of less econometrically problematic measurement error in the y-variable (ad-skip).

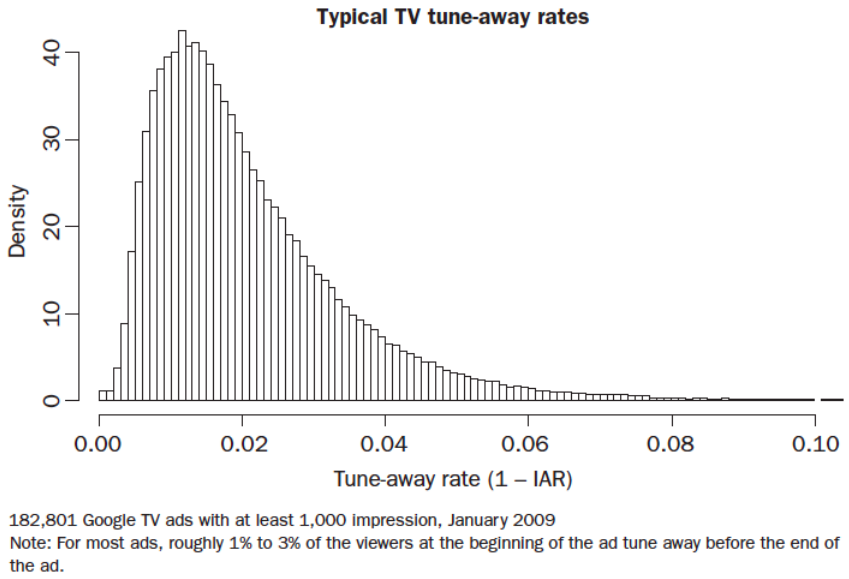


Figure 5: Distribution of Ad Tune-away rates in Dish Network Set-Top Data (reproduced from Interian et al. (2009a), and Zigmod et al. (2009))

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. Analogous results are reported in Zigmond et al. (2009), who report 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 uses only demographics. Using the same individual-level set-top data, Interian et al. (2009b) also report an interesting correlation that the more often viewers have seen an ad over the last month, the less likely they are to tune away.

In our data, we do not observe TV network or show information, but we do observe data on household characteristics. Like Deng, 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. In general, the observed characteristics explain little of the variation in ad exposures and skip rates across households. Later, we will show that historical behavior predicts ad-skipping better than these demographics. 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.

	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.

Table 5: Regression of Household Ad Exposure and Ad Skip Rate on Observed Characteristics

These stylized facts from previous research provide 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 by those consumers of the brands featured in those ads. 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. For instance, the correlation reported in Interian et al. (2009b) can be explained if ads have a positive effect on quantities purchased, and quantities in turn have a positive effect on ad views.

With this exposition of the dataset, we now report on the relationship between purchased quantities and ad consumption in our data, and discuss our strategy to identify complementarities.

3.1 Relating Advertising and Product Demand

First Cut: Cross-sectional Analysis 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. At a minimum, support for such a model requires seeing a positive covariation between quantities and ads in the data. To see if there is preliminary evidence for complementarities, we start by checking whether households who view more advertisements also purchase more on average. We report the joint distribution of total quantity purchased and total category ad consumption at the household level. First, 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.[7] These buckets correspond to households who viewed between 0 and 65 ads, between 65 and 448 ads, and 448+ ads, respectively. Then we non-parametrically estimate the density of purchase quantities for each of these groups separately. Table 6 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.[8]

Based on this cross-sectional analysis, we cannot conclude that advertising *per se* induces this shift out in the purchase density, or that purchase quantities *per se* induce more ad consumption. For example, it could be that larger households are i) more likely to view more advertisements because

[7] Going forward, we restrict our analyses to only include the households who made at least one purchase and were exposed to at least one ad.

[8] 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 ($p = 9 \times 10^{-9}$); we also reject the null for the comparison between the middle two quartiles and the upper quartile of the ad consumption distribution ($p = 1.9 \times 10^{-4}$).

	Purchase Quantity		
Ad Consumption	25th Percentile	50th Percentile	75th Percentile
0<A≤65	7.78	19.54	41.30
65<A≤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 6: Conditional Distribution of Purchase Quantity by Ad Consumption Quartile

they are more likely to own multiple TVs and ii) more likely to buy more because they have a higher demand for the product. Household size 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 effect of complementarities.

Identification of Complementarities Our identification strategy is two-fold. First, we leverage the panel aspect of the data to use only the within-household variation to test for the effect of purchase quantities and ad consumption on each other. The use of the within-variation controls for correlated unobserved heterogeneity. Second, we utilize the 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. In essence, we ask whether all things equal, are more of a product's ads consumed if the household paid low prices in the past? In the panel analysis below, we show that this is the case, as we find that prices affect quantity purchased, and that quantity purchased affects subsequent ad consumption. Under the maintained exclusion restriction, these identify complementarities. Our strategy is analogous to Gentzkow's (2007) contribution on measuring substitution and complementarity between online and offline newspapers. Fox and Lazatti (2014, section 2) provide a formal proof of identification along these lines, showing that access to a variable that shifts the utility from consumption of one good which can be excluded from the utility from consumption of another identifies complementarities in a 2-goods model. Product prices in our data serve the role of this variable.[9] We now present the panel analysis in more detail.

Panel Analysis We use the panel data 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

[9]The success of this identification strategy depends on utilizing extensive within household price variation in the data. For each household in the data and for each brand, we compute the standard deviation in price paid per unit over time as a measure of within household price variability. We find that there is significant *within*-household variation in prices paid over time.

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 how we constructed the price series for these brands.

Given that we have panel data over a long time horizon, we include *household-brand* fixed effects to control for unobserved heterogeneity. Thus, our coefficients are estimated off of within household-brand variation over time rather than across household variation and across brand variation which could have endogeneity concerns as discussed above. In particular, we estimate the following specification.

$$q_{ijt} = \beta_{0ij} + \beta_1 A_{ijt} + \beta_2 p_{ijt} + \beta_3 Time_t + \epsilon_{ijt} \tag{3}$$

where q_{ijt} is daily purchase quantity in equivalent units, A_{ijt} is a cumulative ad-duration variable (defined more precisely in Table 7), and $Time_t$ is a time-trend counting the days since June 14, 2010 (the first day of the data). Table 7 presents the results. Consistent with our cross-sectional findings, the coefficient on cumulative past advertising consumption is positive and statistically significant across all time windows we consider, suggesting that households tend to buy more quantity when they have spent more time watching commercials in the past. 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-consumption over the past 1-4 weeks increases the mean daily quantity demanded by 3.5−6%. For instance, looking at the last 3 rows of Table 7, the mean daily quantity demanded is 7.93 equivalent units. A 1 SD increase in $A_{ijt,7}-$ the ad-consumption over the last 1 week − increases the mean daily quantity demanded by 3.53%.

We now present some robustness checks to these regressions. We report these in Table 8, in which we use our preferred specification in which past ad consumption is defined over the preceding two weeks (Column 2, Table 7). Our first robustness check address a concern there may be unobserved, time-varying shocks driving both purchases and ad consumption that remain even after including household-brand fixed effects.[10] Column 1 adds a dummy for observations in the month of December

[10]In general we would have an endogeneity problem if the firm coordinated prices and advertising quality over time. For example, if low prices lead to high purchase quantities and high-quality ads lead to a higher propensity to watch, we might over-state the relationship between purchase quantity and ad consumption. However, our understanding is that due to the fact that there is little local TV in the Western European country where the data comes from, the manufacturer sets almost all TV advertising at the national level while prices are set by individual retailers at the local level. This suggests that such systematic coordination is unlikely.

	(1) Quantity	(2) Quantity	(3) Quantity	(4) Quantity
$A_{ijt,7}$	0.2254***			
	(0.0337)			
$A_{ijt,14}$		0.1549***		
		(0.0210)		
$A_{ijt,21}$			0.1194***	
			(0.0160)	
$A_{ijt,28}$				0.1241***
				(0.0139)
Price Per Unit	-14.931***	-14.279***	-13.012***	-13.386***
	(2.6402)	(2.6848)	(2.6924)	(2.7727)
Time Trend	0.0014***	0.0011***	0.0008***	0.0001
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Observations	29,590,182	29,218,979	28,847,776	28,476,573
HH-Brand FE	Y	Y	Y	Y
Ave Quantity	7.93	7.95	7.96	8.02
Effect of +1 SD Ads	0.28	0.34	0.37	0.48
Percent of Mean D.V.	3.53%	4.28%	4.65%	5.99%
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 7: Regression of Daily Purchase Quantity on Cumulative Ad Consumption

	(1) Quantity	(2) Quantity
$A_{ijt,14}$	0.1943***	2.5047***
	(0.0211)	(0.5053)
Price Per Unit	-13.5245***	-343.43***
	(2.6878)	(21.621)
December	3.4976***	
	(0.1471)	
Time Trend	-0.0001	0.0270***
	(0.0003)	(0.0058)
Observations	29,218,979	883,736
HH-Brand FE	Y	Y
Ave Quantity	7.95	262.74
Effect of +1 SD Ads	0.43	5.82
Percent of Mean D.V.	5.41%	2.21%
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.

Table 8: Robustness Checks of Regression of Daily Purchase Quantity on Cumulative Ad Consumption

to check if the results are robust to a story that there may be increased demand for the product around the holidays and, at the same, the intensity of advertising may be higher and ad content may be more engaging during those times. We find our results are robust to this additional control. Another story along these lines could be that 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 check whether our results are driven by such a scenario, we re-estimate the same model, restricting the data to days in which a household purchased at least one brand. Again, we continue to estimate a positive relationship between purchase quantity and cumulative ad consumption.

We also estimate the analogous model as above on the ad side, but now treating current advertising consumption as the dependent variable and cumulative product consumption of the related brand as a regressor. As noted before, whether or not a consumer is exposed to an ad is determined by firms' supply of advertisements and consumers' show-preferences. However, conditional on being exposed to an ad, consumers have agency over how much of the ad to watch. Hence, we believe the percentage of the ad watched conditional on exposure is a better metric of advertising demand. In the regressions below, we treat the percent watched (range: 0-1) of the r^{th} ad exposure for brand j watched by household i in day t on household i's cumulative past consumption of brand j, a_{ijrt}, as

	(1) Percent Ad Watched Conditional on Exposure	(2) Percent Ad Watched Unconditional on Exposure
$Q_{ijt,14}$	3.83e-07*	6.01e-07***
	(2.10e-07)	(1.53e-07)
Time Trend	5.47e-05***	9.08e-05***
	(1.75e-06)	(2.20e-06)
Observations	1,436,400	29,778,304
HH-Brand FE	N	Y
HH FE	Y	N
Brand FE	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. Column 1 is estimated conditional on an exposure. In column 2, the dependent variable is recorded as a 0 for days in which no ads were viewed. $Q_{ijt,14}$ records the cumulative package volume household i purchased of brand j in the 14 days preceding day t. Robust standard errors clustered at the household level.

Table 9: Regressions of Ad Consumption on Cumulative Quantity

the dependent variable. Each row in the regression is a household-brand-day-exposure combination. This specification is summarized in equation 4.

$$a_{ijrt} = \theta_{0i} + \theta_{0j} + \theta_1 Q_{ijt} + \theta_2 Time_t + \epsilon_{ijrt} \tag{4}$$

Here, Q_{ijt} is the cumulative quantity variable, and $Time_t$ is a similar time-trend counting the days since June 14, 2010 (the first day of the data). Table 9 reports the results. Column 1 reports on the effect of cumulative quantity purchased over the past 2 weeks on the percentage of an ad viewed by a household, conditional on exposure. Since we are doing everything conditional on exposure, we lose power and can include only brand and household fixed effects as opposed to brand-household fixed effects. Looking at column 1, past quantity is seen to have a positive and significant effect on ad-skipping (we report marginal effects below). For completeness, we also run 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 2, 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.

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
$Q_{ijt,14}$	3.83e-07*	3.85e-07*	4.34e-07	7.67e-07	6.06e-07	2.73e-06	6.54e-06
	(2.10e-07)	(2.19e-07)	(4.12e-07)	(9.52e-07)	(2.36e-06)	(3.87e-06)	(6.21e-06)
Time Trend	5.47e-05***	5.67e-05***	8.96e-05***	0.000142***	0.000180***	0.000227***	0.000283***
	(1.75e-06)	(1.80e-06)	(3.57e-06)	(7.72e-06)	(1.38e-05)	(2.52e-05)	(3.76e-05)
Observations	1,436,400	1,369,030	558,595	159,242	57,764	23,522	12,327
R-squared	0.039	0.038	0.036	0.049	0.064	0.083	0.096
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
Marginal Effect of an Additional Purchase of Various Quantities on Expected Percentage Watched							
Mean	0.08%	0.08%	0.09%	0.16%	0.12%	0.55%	1.31%
25th Percentile	0.03%	0.04%	0.04%	0.07%	0.06%	0.26%	0.60%
50th Percentile	0.06%	0.06%	0.07%	0.11%	0.09%	0.40%	0.96%
75th Percentile	0.09%	0.09%	0.10%	0.17%	0.14%	0.62%	1.48%
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. $Q_{ijt,14}$ records the cumulative package volume household i purchased of brand j in the 14 days preceding day t. 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.

Table 10: Regressions of Daily Ad Consumption on Cumulative Quantity for Households with Different Skip Rates

To explore the heterogeneity in advertising consumption effects, we repeat the same regression separately for households of different observed ad-skip rates. Table 10 repeats the regression from Column 1 of Table 9 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. 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 and brand fixed effects. For ease of comparison, Column 1 in Table 10 repeats the results from Column 1 of Table 9. Although we lose power when focusing on only households that have high skip-rates, the effects 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 10, 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. Then, 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 $> 25\%$. Denote the mean, 25th, 50th and 75th percentile of the quantity purchased by households in that subgroup conditional on purchase as $(\bar{q}, q_{.25}, q_{.5}, q_{.75})$ respectively. 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 $q_{.5}$, households in that subgroup are likely to watch 0.96% more of an ad, conditional on exposure; and if the quantity purchased over the previous two weeks is increased by $q_{.75}$, households in that subgroup are likely to watch 1.48% more of an ad, conditional on exposure.

The above regressions pooled data across households including household and brand specific intercepts, but restricted the slope coefficients to be the same. Different households may have different sensitivities in how their purchase quantity is related to their ad consumption, and the un-modeled slope heterogeneity may be a source of spurious within-household correlation. To address this concern, we also run the above regressions separately for each household, in essence allowing the coefficients on cumulative quantity and cumulative ad time to differ for each household. For each household, we separately estimate the following regressions,

$$q_{jt} = \beta_{0j} + \beta_1 A_{jt,14} + \beta_2 p_{jt} + \beta_3 Time_t + \epsilon_{jt} \tag{5}$$

$$a_{jrt} = \theta_0 + \theta_1 Q_{jt,14} + \theta_2 Time_t + \epsilon_{jrt} \tag{6}$$

Coeff	Brand FE	-	n.s.	+
β_{1i}	Y	3%	87%	10%
θ_{1i}	Y	10%	74%	16%

Table 11: Significance of Household Specific Coefficients

where the index i for household is suppressed for brevity. The implicit assumption here is that parameters are time-invariant for a given household. Table 11 records the percentage of coefficients that are positive and significant, negative and significant, and not statistically significant at the 5% level, and Figure 6 plots the empirical CDFs of the two sets of coefficients across households. Though we lose power especially for households with few purchase and ad exposure observations, we see that the cross effects of quantity and advertising are positive for a large subset.

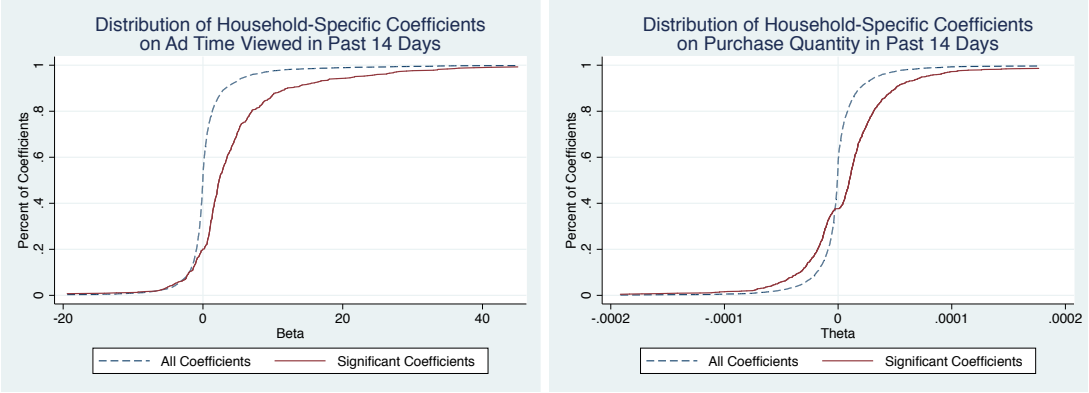


Figure 6: Histogram of Household-Specific $\hat{\beta}_{1i}$ and $\hat{\theta}_{1i}$ from Model with Brand Fixed Effects

Summary To summarize our results so far, we find that households tend to buy more when they have been exposed to more advertising in the past, and also tend to watch more advertising on the margin when they have purchased more in the past. These patterns do not seem to be driven by correlated tastes that drive both product purchase and ad viewing behavior. The fact that past purchase quantities seem to systematically explain the within-household variation in ad consumption rates also suggests that advertising is not simply playing a reminder role, and that complementarities are the more likely explanation.

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 do have agency over whether to watch a whole commercial or decide 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), we recognize that it is unlikely that consumers make a simultaneous decision of which goods to buy and how much of each advertisement to watch. It is more likely that consumers make these decisions sequentially during the day. The sequential model of daily decisions we use is depicted in Figure 7. 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 8 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.

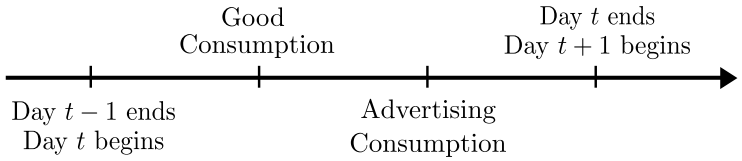


Figure 7: Timing of Consumption Decisions

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 subse-

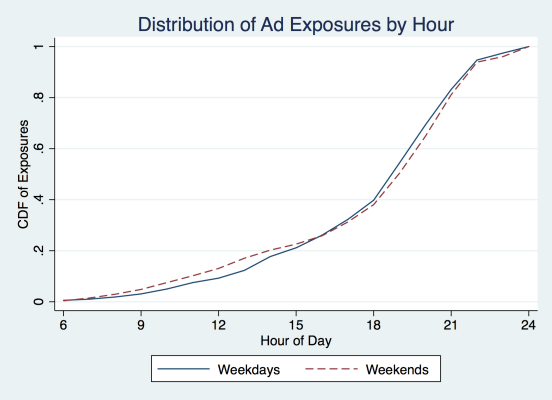


Figure 8: Ad Exposures by Time of Day

quent 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, when we model ad-skipping, we model only a binary variable of the consumer's action of seeing the ad fully versus skipping it towards another channel or activity. On the ad-side, we model the skip or not decision, and not the continuous decision of *how much* of an ad to watch because we believe the latter indicator would be a noisier measure to assess underlying consumer preferences. Rather, the key preference over advertisements is revealed in the decision to skip the ad or not. We think this modeling choice is more consistent with the actual trade-off consumers make.[11] 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, 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

[11] In Appendix B, 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.

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). 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:

$$\max_{\boldsymbol{x}_t} \quad U^G\left(\boldsymbol{x}_t \middle| A_{t-1}\right) \quad s.t.\ \boldsymbol{p}_t^G \cdot \boldsymbol{x}_t \leq E_t^G \tag{7}$$

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, 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, 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:

$$U^G\left(x_{0t}...x_{Jt} \middle| A_{t-1}\right) = U_0^G\left(\cdot\right) + U_1^G\left(\cdot\right) \tag{8}$$

where

$$U_0^G\left(x_{0t}\right) = e^{\gamma_0} x_{0t}.e^{\varepsilon_{0t}^G} \tag{9}$$

$$U_1^G\left(x_{1t}...x_{Jt} \middle| A_{t-1}\right) = \sum_{j=1}^{J} \exp\left[\gamma_j + \beta \log\left(1+A_{jt-1}\right)\right]\left(1+x_{jt}\right)^{\delta_j}.\mathrm{e}^{\varepsilon_{jt}^G} \tag{10}$$

where $0<\delta_j<1$, is a satiation parameter. Above, $\varepsilon_{0t}^G...\varepsilon_{Jt}^G$ are i.i.d. stochastic shocks that are known to the consumer but not to the econometrician. The utility function is additively separable across brands, which are assumed to be substitutes. However, complementarities between consumption and advertising of a given brand are allowed. The implied marginal utility of consumption is,

$$\frac{\partial U^G}{\partial x_{jt}} = \left(\frac{\delta_j \exp\left[\gamma_j + \beta \log\left(1+A_{jt-1}\right)\right]}{\left(1+x_{jt}\right)^{1-\delta_j}}\right).\mathrm{e}^{\varepsilon_{jt}^G} \tag{11}$$

The marginal utility of consuming x_{jt} depends on the quantity purchased 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,

$$\frac{\partial}{\partial A_{jt-1}}\frac{\partial U^G}{\partial x_{jt}} = \left(\frac{\beta\delta_j \exp\left[\gamma_j + \beta \log\left(1+A_{jt-1}\right)\right]}{\left(1+x_{jt}\right)^{1-\delta_j}\left(1+A_{jt-1}\right)}\right).\mathrm{e}^{\varepsilon_{jt}^G} \tag{12}$$

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,

$$\max_{\boldsymbol{x}_t \geq 0, \lambda_t \geq 0} \mathbb{L}^G(x_t, \lambda) = U^G(\boldsymbol{x}_t | A_{jt-1}) + \lambda_t \left(E_t^G - \sum_j \boldsymbol{p}_{jt} \cdot \boldsymbol{x}_{jt} \right) \tag{13}$$

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 of the others by a household in 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,

$$\frac{\partial \mathbb{L}^G}{\partial x_{jt}} = 0, \quad x_{jt}^* > 0,\, j \leq K \tag{14}$$

$$\frac{\partial \mathbb{L}^G}{\partial x_{jt}} \leq 0, \quad x_{jt}^* = 0,\, j > K \tag{15}$$

$$x_{0t}, x_{jt}, \lambda_t \geq 0, \quad j \in J \tag{16}$$

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., Chintagunta and Nair 2011), it follows that at the optimum,

$$V_{0t} - V_{jt} = \varepsilon_{jt} - \varepsilon_{0t},\, x_{jt}^* > 0,\, j \leq K \tag{17}$$

$$V_{0t} - V_{jt} \geq \varepsilon_{jt} - \varepsilon_{0t},\, x_{jt}^* = 0,\, j > K \tag{18}$$

where,

$$V_{0t} = \gamma_0 \tag{19}$$

$$V_{jt} = \gamma_j + \beta \log(1 + A_{jt-1}) + \log(\delta_j) - (1 - \delta_j) \log(1 + x_{jt}) - \log(p_{jt}) \tag{20}$$

We can now derive the implied joint probabilities of purchase and density of quantities by integrating the random utility errors, $\varepsilon_{0t}^G \ldots \varepsilon_{Jt}^G$, over the implied KKT conditions in Equations (17 and 18). Assuming $\varepsilon_{0t}^G \ldots \varepsilon_{Jt}^G$ follow a Type-1 extreme-value distribution, the mixed discrete-continuous density of the outcomes for an agent at time t conditional on a vector of price and past advertising becomes,

$$l(x_{0t}^*, \ldots, x_{Kt}^*, 0, \ldots, 0 | p_t, A_{t-1}) = (K-1)! . \frac{\Pi_{j=0}^K \exp(V_{jt})}{\left(\Sigma_{j=0}^J \exp(V_{jt})\right)^K} . |\mathcal{J}| \tag{21}$$

where $V_{.t}$ are as defined in Equation (20) and $|\mathcal{J}|$ 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 elements of $\mathcal{J}$ are,

$$[\mathcal{J}]_{l,k} = \frac{\partial\left(V_{0t} - V_{lt} + \varepsilon_{0t}\right)}{\partial x^*_{kt}} = \mathbf{1}\left(l = k\right)\frac{1-\delta_k}{1+x^*_{kt}},\ l,k = 1..K \tag{22}$$

and its determinant is $\left[\Pi_{k=0}^{K} f_{kt}\right]$, where $f_{kt} = \frac{1-\delta_k}{1+x_{kt}}$, $k = 1,..,K$.

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,

$$\max_{w_{st}\in\mathbb{R}^+, a_{sjt}\in(0,1)} \quad U^A\left(w_{st}, a_{sjt}|\, X_{jt}\right) \quad s.t.\ \ w_{st} + \tau a_{sjt} \leq T \tag{23}$$

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

$$U^A\left(w_{st}, a_{sjt}|\, X_{jt}\right) = U_0^A\left(w_{st}\right) + U_1^A\left(a_{sjt}|\, X_{jt}\right) \tag{24}$$

where,

$$\begin{aligned} U_0^A\left(w_{st}\right) &= e^{\alpha_0}.w_{st}.e^{\varepsilon^A_{s0t}} \\ U_1^A\left(a_{sjt}|\, X_{jt}\right) &= a_{sjt}.\exp\left[\alpha_j + \theta\log\left(1+X_{jt}\right)\right].\mathrm{e}^{\varepsilon^A_{sjt}} \end{aligned}$$

and $\varepsilon^A_{s0t}, \varepsilon^A_{sjt}$ 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 (23), $w_{st} + \tau \times 0 = T$, so $w_{st} = T$ when $a_{sjt} = 0$. We can now obtain the conditional indirect utility from skipping by evaluating the direct utility (23) at $w_{st} = T$ and $a_{sjt} = 0$,

$$\begin{aligned} V^A\left(a_{sjt} = 0|X_{jt}\right) &= U_0^A\left(w_{st} = T\right) + U_1^A\left(a_{sjt} = 0|\, X_{jt}\right) \\ &= e^{\alpha_0}.T.e^{\varepsilon^A_{s0t}} \end{aligned}$$

If the consumer decides to watch ad-exposure s, $a_{sjt} = 1$. Substituting again for a_{sjt} into the time-

constraint in Equation (23), $w_{st} + \tau \times 1 = T$, 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 (23) at $w_{st} = T - \tau$ and $a_{sjt} = 1$,

$$\begin{aligned} V^A\left(a_{sjt}=1|X_{jt}\right) &= U_0^A\left(w_{st}=T-\tau\right)+U_1^A\left(a_{sjt}=1|\,X_{jt}\right) \\ &= e^{\alpha_0}.\left(T-\tau\right).e^{\varepsilon^A_{s0t}}+\exp\left[\alpha_j+\theta\log\left(1+X_{jt}\right)\right].\mathrm{e}^{\varepsilon^A_{sjt}} \end{aligned}$$

The consumer will not skip the ad if $V^A\left(a_{sjt}=1|\,X_{jt}\right) > V^A\left(a_{sjt}=0|\,X_{jt}\right)$. 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,

$$\mathcal{W}_{sjt} - \mathcal{W}_{s0t} \geq \varepsilon^A_{s0t} - \varepsilon^A_{sjt} \tag{25}$$

The implied probability of watching an advertisement is a logit,[12]

$$Pr\left(a_{sjt}=1\right) = \frac{\exp\left(\mathcal{W}_{sjt}\right)}{\exp\left(\mathcal{W}_{s0t}\right)+\exp\left(\mathcal{W}_{sjt}\right)} \tag{26}$$

Evidence of complementarities for the agent can be found if,

$$\frac{\partial}{\partial X_{jt}}\left[U_1^A\left(a_{sjt}=1|\,X_{jt}\right)-U_1^A\left(a_{sjt}=0|\,X_{jt}\right)\right] = \frac{\theta}{1+X_{jt}}\exp\left[\alpha_j+\theta\log\left(1+X_{jt}\right)\right].\mathrm{e}^{\varepsilon^A_{sjt}} \geq 0 \tag{27}$$

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 (21). 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.,

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

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

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

[12] τ is not identified separately from the intercept and is absorbed into α_0.

where,

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

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 in 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,

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

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:

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

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,

$$Pr\left(b_{1t}, ..., b_{zt} |\, z_t, X_t\right) = \prod_{s=1}^{z_t} \prod_{j=1}^{J} \phi_j^{1(b_{st}=j)} \tag{28}$$

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,

$$l\left(a_{0t}^*, ..., a_{zt}^* |\, z_t, b_{1t}, ..., b_{zt}, X_{t-1}\right) = \prod_{s=1}^{z_t} \left(\frac{\exp\left(\mathcal{W}_{sjt}\right)}{\exp\left(\mathcal{W}_{s0t}\right) + \exp\left(\mathcal{W}_{sjt}\right)}\right)^{a_{sjt}} \left(\frac{\exp\left(\mathcal{W}_{s0t}\right)}{\exp\left(\mathcal{W}_{s0t}\right) + \exp\left(\mathcal{W}_{sjt}\right)}\right)^{1-a_{sjt}} \tag{29}$$

To reiterate, our ad-utility parameters $(\boldsymbol{\alpha}, \theta)$ are identified purely off the conditional skip/no-skip decision (equation 29), 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. Accordingly, we restrict Σ to be a block-diagonal matrix with the exception of non-zero covariance terms that we estimate on the $\beta \times \theta$ 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 12. The estimates of the remaining auxiliary parameters (λ, π, ϕ_j) are included in a table in Appendix C. Looking first at the estimates for the product consumption model in Table 12, 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 relatively large magnitude of the σ^γ parameters indicates significant heterogeneity in purchase frequency across households. The δ parameters help pin down purchase quantities by allowing for satiation. The fact that δ_6 is smaller than δ_j is consistent

Brand	$\bar{\gamma}_{ij}$	σ^{γ}_{ij}	$\bar{\alpha}_{ij}$	σ^{α}_{ij}
1	-5.5126	1.4344	-	-
	(0.0455)	(0.0396)	-	-
2	-5.5236	1.6692	3.1545	0.6845
	(0.477)	(0.0463)	(0.0196)	(0.0285)
3	-2.6025	1.1760	3.3524	0.4290
	(0.0225)	(0.0206)	(0.0184)	(0.0455)
4	-4.4487	1.4844	3.3087	0.6020
	(0.0372)	(0.0335)	(0.0172)	(0.0248)
5	-4.9791	1.3104	2.7376	0.7714
	(0.0390)	(0.0293)	(0.0189)	(0.0285)
6	-3.6492	1.5969	-	-
	(0.0362)	(0.0276)	-	-
7	-4.6806	1.2513	2.7002	0.7520
	(0.0361)	(0.0328)	(0.0149)	(0.0173)
8	-5.5188	2.0069	2.8420	0.7858
	(0.0516)	(0.0426)	(0.0161)	(0.0180)
9	-5.1687	1.4472	2.5662	0.8517
	(0.0417)	(0.0332)	(0.0171)	(0.0207)
10	-5.5984	1.6468	2.6739	0.7176
	(0.0512)	(0.0442)	(0.0168)	(0.0222)
11	-3.3804	1.3692	2.9662	0.7394
	(0.0270)	(0.0228)	(0.0190)	(0.0250)

β	σ^{β}	θ	σ^{θ}	ρ
0.1111	0.0130	0.0053	0.0132	0.0303
(0.0093)	(0.0004)	(0.0031)	(0.0002)	(0.0001)

δ_j	δ_6
0.8321	0.8006
(0.0007)	(0.0026)

LL = -1,194,298
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 $\beta \times \theta$ off-diagonals. Model estimated via maximum simulated likelihood using 1,000 draws of Θ_i per household to integrate over the implied random effects distribution. To enforce support on $(0, 1)$, the satiation parameters $\boldsymbol{\delta}$ are transformed as $\boldsymbol{\psi} = \log \frac{\boldsymbol{\delta}}{1-\boldsymbol{\delta}}$, and an unrestricted search over $\boldsymbol{\psi}$ implemented in the likelihood routine. Standard errors for $\boldsymbol{\delta}$ reported using the delta method. We estimate two separate δ-s, one for brand 6, and another which is common for all other brands.

Table 12: Parameter Estimates for the Joint Model Estimated on the Full Sample

with the fact that brand 6 is observed to be purchased in smaller quantities than the other brands. The complementary 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. The standard deviation σ^{θ} is relatively large so there is a reasonably large chance of having a negative θ. Finally, the estimated correlation between β and θ is captured by ρ and is estimated to be positive but small.

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,

$$\hat{\bar{\Theta}}_i = \int \Theta_i \frac{\mathcal{L}_i(\Theta_i|\hat{\bar{\Theta}}, \hat{\Sigma})}{\mathcal{L}_i(\hat{\bar{\Theta}}, \hat{\Sigma})} d\Phi(\Theta_i|\hat{\bar{\Theta}}, \hat{\Sigma}) \tag{30}$$

We use this "approximate Bayesian" approach (Allenby and Rossi 1999; 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 9 and 10. Table 14 at the end of this section summarizes the analogous results for all the advertised brands in the data.

Consider brand 4. We simulate the increase in exposures for the brand as follows. We first calculate

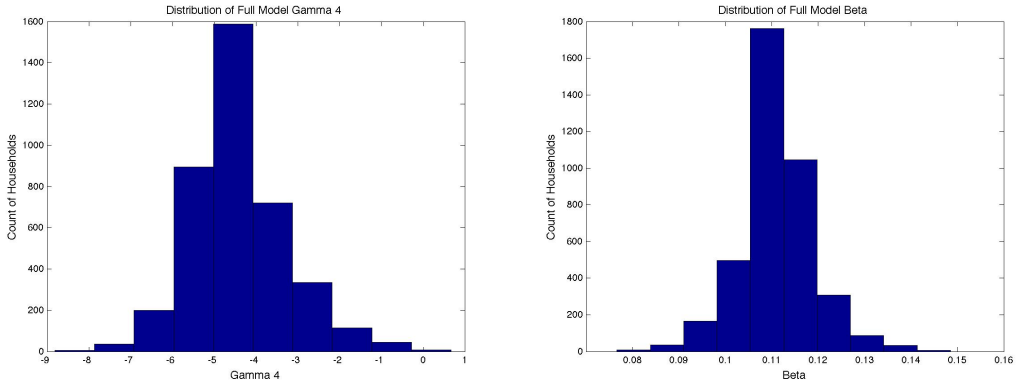


Figure 9: Household Specific Estimates of γ_4 and β

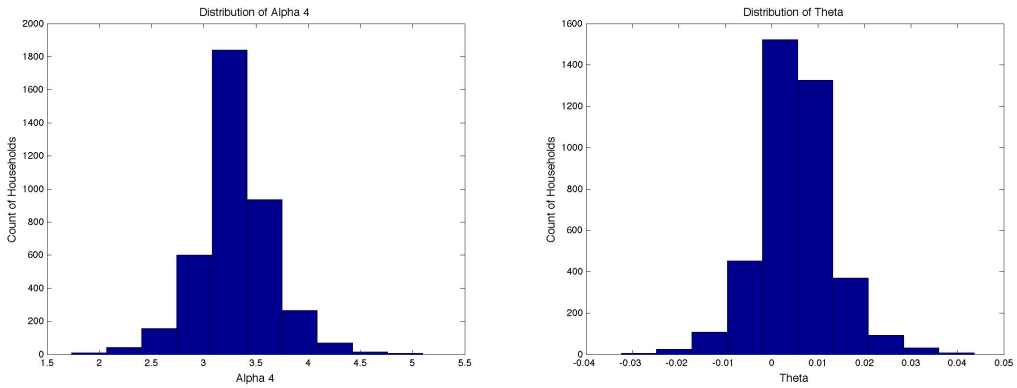


Figure 10: Household Specific Estimates of α_4 and θ

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 3,269 households who viewed at least one advertisement for brand 4 during these 3 months.

Endogenous Non-Compliance: Ad Skipping Figure 11 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.

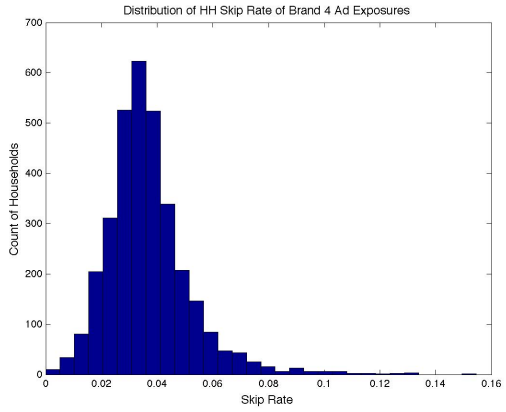


Figure 11: Distribution of Simulated Household Ad Skip Rates of Brand 4 Exposures

Figure 12 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 11 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 25.6 to 27.9 ads across households. This reflects the differential uptake of treatment across households.

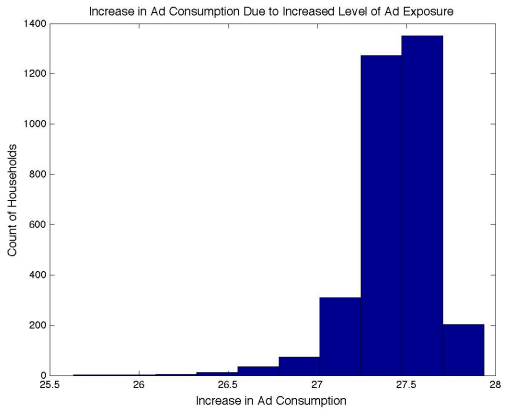


Figure 12: Increase in Consumption of Brand 4 Ads Resulting from an Increase in Number of Ad Exposures

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 31, 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 10. We sum this utility across the last three months of the data to obtain the total utility from consumption. We report a "welfare elasticity" calculated as the percent increase in total utility from consumption over the three months in the simulation divided by the percent increase in ad consumption.

$$\eta_{ij}^{q} = \frac{\Delta Q_{ij}/Q_{ij}}{\Delta A_{ij}/A_{ij}}, \eta_{ij}^{u} = \frac{\Delta U_{i}/U_{i}}{\Delta A_{ij}/A_{ij}} \tag{31}$$

These elasticities can be thought of as long-run effects of advertising. Figure 13 shows the distributions of model-predicted demand and welfare elasticities with respect to advertising across households. The mean advertising elasticity of demand is 0.3501 and the mean advertising elasticity of welfare is 0.0261.[13] There is significant heterogeneity across households.

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

[13] Across brands, we estimate advertising elasticities of demand ranging from 0.11 to 0.35. These estimates are comparable to the advertising elasticity of demand of 0.15 that Ackerberg (2001) estimates in his analysis of TV commercials and demand for yogurt.

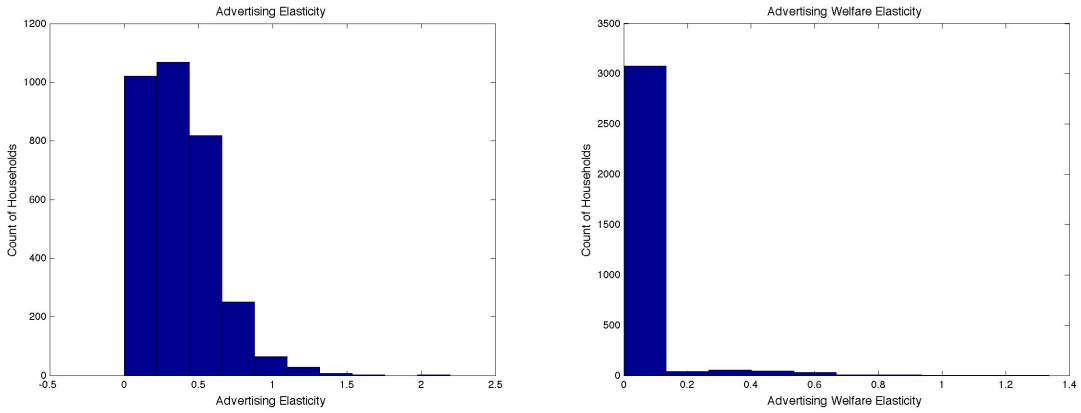


Figure 13: Advertising Elasticities

	(1)	(2)
	% Δ in Purchase Quantity$_i$	% Δ in Utility$_i$
ΔA$_i$	0.0960**	0.0311**
	(0.0452)	(0.0144)
Constant	-2.1363	-0.8149**
	(1.2391)	(0.3948)
Observations	3,266	3,269
Mean Dep Var.	0.4971	0.0372
Standard errors in parentheses		
*** p<0.01, ** p<0.05, * p<0.1		

Note: Regression estimated at the household level. The independent variable ΔA$_i$ records each household's model-predicted increase in ad consumption between the two simulations. The dependent variables record the model-predicted percent increase in purchase incidence and purchase quantity between the two simulations.

Table 13: Regression of Percent Increase in Purchase Outcomes on Increase in Ad Consumption

these outcome variables of interest. We regress the percent increase in predicted quantity and utility on the model-predicted incremental ad time watched. The results are in Table 13. 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 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 and welfare will likely change in response to the campaign.

These simulations carried out with our structural model illustrate the existence of endogenous non-compliance with advertising and the resulting implications for measuring advertising effectiveness. In the next section we explore how our model's predictions compare to the predictions of a model in which advertising is not endogenized.

		Full Model			Traditional Model	
Brand	Δ Ad Exposures	Δ Ad Consumption	Demand Elasticity	Welfare Elasticity	Demand Elasticity	Welfare Elasticity
2	8	7.8061	0.1760	0.0057	0.1616	0.0059
		(0.0934)	(0.2379)	(0.0453)	(0.2167)	(0.0480)
3	25	24.5356	0.2747	0.1102	0.2532	0.0996
		(0.1356)	(0.2749)	(0.2133)	(0.2528)	(0.1933)
4	28	27.4366	0.3501	0.0261	0.3264	0.0251
		(0.2197)	(0.2702)	(0.1101)	(0.5462)	(0.1040)
5	9	8.6707	0.1898	0.0116	0.1732	0.0107
		(0.1654)	(0.2548)	(0.0826)	(0.2310)	(0.0709)
7	11	10.5916	0.2234	0.0142	0.2082	0.0133
		(0.1938)	(0.2764)	(0.0921)	(0.2559)	(0.0836)
8	0	-	-	-	-	-
		-	-	-	-	-
9	12	11.4799	0.2279	0.0125	0.2109	0.0110
		(0.2942)	(0.2685)	(0.0816)	(0.2506)	(0.0758)
10	6	5.7734	0.1120	0.0052	0.1022	0.0040
		(0.1098)	(0.2096)	(0.0490)	(0.1809)	(0.0375)
11	10	9.7110	0.1649	0.0318	0.1519	0.0290
		(0.1547)	(0.2204)	(0.1109)	(0.2031)	(0.0996)

Note: Table reports the mean 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.

Table 14: Summary of Simulation Results by Brand

6 Comparison to the Traditional Model

We call a model in which advertising is not endogenized the "traditional" model. We think of the traditional model as the product demand model a researcher might estimate if he only observed data on ad exposure and not ad consumption. In particular, in this model, the stock of advertising A_{t-1} would simply record the integer number of exposures a household viewed over the previous 14 days. This is in contrast to our model in which the stock of advertising records the sum-total of the fraction of each ad exposure consumed over the previous 14 days. Further, the "traditional" model will not have a likelihood related to ad consumption. To do our comparison, we start by fitting a "traditional" model to the same data via maximum simulated likelihood. Table 15 reports the estimates for this model.

How would dropping the ad consumption component bias the parameters of this "traditional" model relative to the "full" model? To build some intuition, consider a linear set-up in which quantity Q is linearly related to ad consumption A :

$$Q = A\beta + v$$

where v represent unobservables that drive Q. The traditional model replaces ad consumption, A, by ad exposures, $\tilde{A}$. We can write $\tilde{A} = A + u$, where u is the measurement error associated with using exposures instead of consumption. u can also be thought of as the skip-rate (so, exposure is $=$ ad consumption $+$ skip-rate). The key aspect to note is that the measurement error is non-classical because $\text{corr}\,(A, u) \neq 0$ (because the set of consumers who decide to skip an ad may also be more likely to be exposed to it) and $\text{corr}\,(u, v) \neq 0$ (because the propensity to skip an ad $-$ embedded in the skip-rate u $-$ is related to the propensity to consume more units of the product $-$ embedded in the unobservables driving quantities, v). Substituting for $A = \tilde{A} - u$, and plugging into the Q-equation, the traditional model estimated using $\tilde{A}$ is, $Q = \tilde{A}\beta + (v - u\beta)$. So, asymptotically what we obtain from this model is,

$$\text{plim}\,\hat{\beta} = \text{plim}\left(\tilde{A}'\tilde{A}\right)^{-1}\left(\tilde{A}'Q\right) = \beta + \text{plim}\left(\tilde{A}'\tilde{A}\right)^{-1}\tilde{A}'v - \text{plim}\,\beta\left(\tilde{A}'\tilde{A}\right)^{-1}\tilde{A}'u$$

Both the last two terms are not zero, and depending on their relative magnitude, the bias induced in measuring the effect of ads in the "traditional" model can go either way depending on patterns of taste covariation in the population. When the effect of ads on quantities is biased upward, the standard treatment overstates the welfare effects of ads, and when the effect of advertising is biased downward, the standard treatment understates the effects of ads.

We start by documenting the differences in parameter estimates between the two models. We obtain household-level parameter estimates for the traditional model using Equation 30 applied to the

Brand	$\bar{\gamma}_{ij}$	σ_{ij}^{γ}
1	-5.4777	1.5069
	(0.0481)	(0.0416)
2	-5.4535	1.7194
	(0.0462)	(0.0439)
3	-2.6165	1.1713
	(0.0220)	(0.0203)
4	-4.5873	1.6846
	(0.0401)	(0.0335)
5	-4.8891	1.4152
	(0.0372)	(0.0302)
6	-3.6004	1.7984
	(0.0369)	(0.0376)
7	-4.5932	1.3226
	(0.0343)	(0.0310)
8	-5.4436	2.0573
	(0.0504)	(0.0387)
9	-5.2078	1.7096
	(0.0455)	(0.0398)
10	-5.6404	1.7618
	(0.0519)	(0.0403)
11	-3.3901	1.4166
	(0.0259)	(0.0271)

β	σ^{β}
0.1041	0.0175
(0.0086)	(0.0262)

δ_j	δ_6
0.8314	0.7980
(0.0006)	(0.0024)

LL = -931,602
No. of households = 4,221
No. of observations = 2,325,771

Table 15: Parameter Estimates for the Traditional Model Estimated on the Full Sample

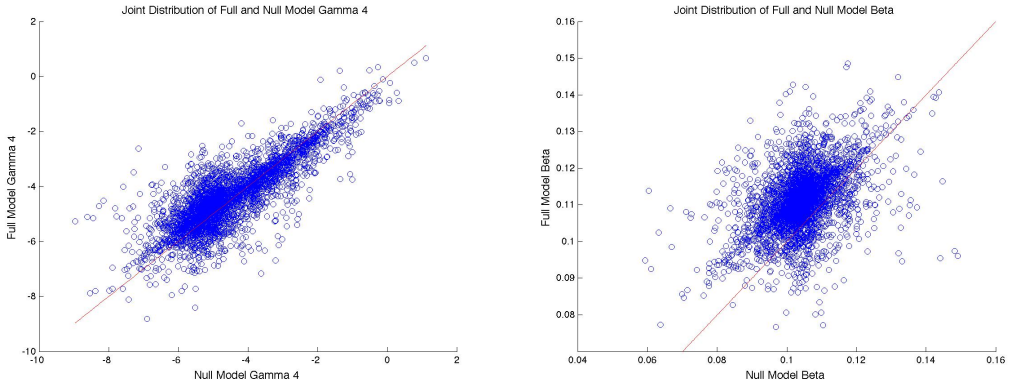


Figure 14: Household Specific γ_4 and β Estimates for the Full and Traditional Models

estimates of this model reported in Table 15, and compare them to the household-level estimates for the full model. Figure 14 plots the full model's estimates against the traditional model's estimates of γ_4 and β. Recall that γ_j is a brand intercept and β is the parameter on advertising in the consumption model. The complementarities parameter $\hat{\beta}_{4i}^F$ in the full model is larger than $\hat{\beta}_{4i}^T$ in the traditional model for about 86% households.[14]

Next, just as we did with the full model, we use these household-level estimates to simulate purchase outcomes over the last quarter of 2011 at the observed level of ad exposure and an increased level of ad exposure. We use the same error draws in these simulations that we used in the full model simulations. Figure 15 plots the increase in ad consumption under the full and traditional models. All points have the same x-coordinate because the traditional model predicts that all households experience the same increase of 28 ads in our simulation. In contrast, as discussed in the previous section, the full model predicts that some households experience a larger increase than others because households can skip ads and there is heterogeneity in skip rates across households. As before, this reflects endogenous take-up of ads.

We also calculate elasticities using the traditional model and compare these elasticities to the full model elasticities.[15] The mean advertising demand elasticity for brand 4 in the traditional model is 0.3264. This is about 7% lower than the mean elasticity predicted using the full model. The mean advertising welfare elasticity for brand 4 is 0.0251, which is 4% lower than the mean elasticity predicted using the full model. Figure 16 plots the histogram of the difference between the full model and traditional model brand 4 demand and welfare elasticities. On average the full model predicts

[14] We tested whether the full model's estimate of $\hat{\beta}_{4i}$ is relatively larger for households who have high skip rates. We regress the ratio of $\hat{\beta}_i^F/\hat{\beta}_i^T$ on observed household skip rates. We found the coefficient on observed skip rates is positive, but not statistically significant.

[15] The elasticities for the traditional model are calculated using the percent increase in exposures. For internal consistency, the elasticities for the full model presented in Table 14 are calculated using the percent increase in ad consumption. We also calculated but do not show the elasticities for the full model using percent increase in ad exposures and the results are almost identical.

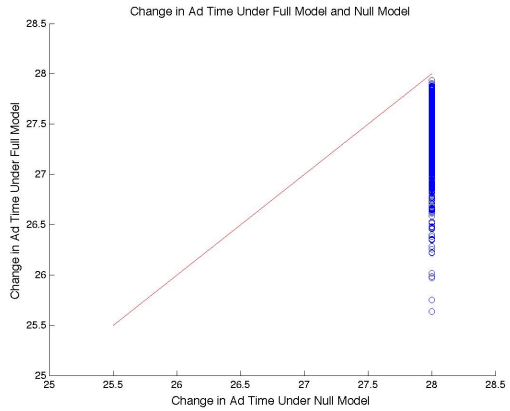


Figure 15: Change in Ad Time Under Full Model and Traditional Model

higher elasticities, but there are households for whom the traditional model predicts a higher elasticity. Table 14 reports the mean and standard deviation of the distribution of household-level elasticities for all other brands. The results show that the traditional model on average underestimates demand and welfare elasticities. We are still working on building clear intuition for these comparative differences.

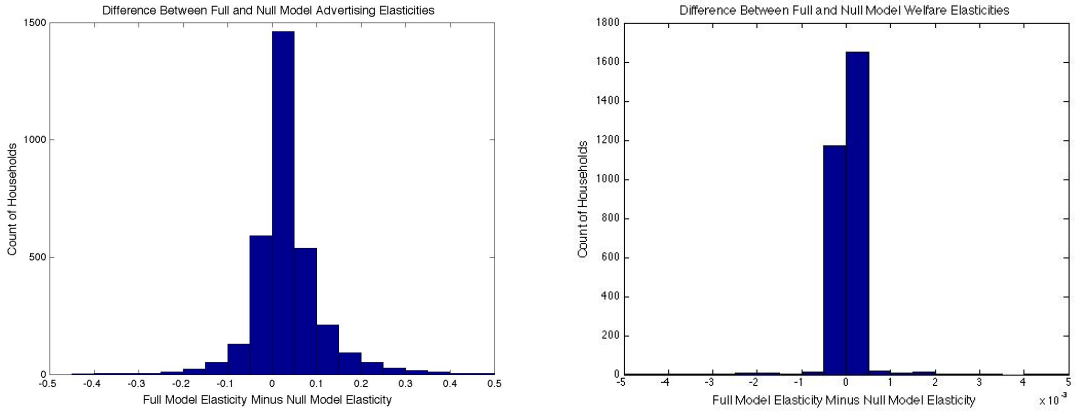


Figure 16: Difference Between Full Model and Traditional Model Demand and Welfare Elasticities

Discussion These simulations underscore the fact that failing to account for endogenous non-compliance with treatment may lead to mis-measurement of treatment effects. Comparing the results from our full model to the traditional model in which advertising is not endogenized, we find that failing to endogenize advertising on average leads to under-estimation of advertising elasticities. To close the paper, we explore another useful consequence of endogenizing advertising consumption, namely that it leads to a natural way of incorporating the quality of advertisements into the model.

7 Quality Simulation

In this counterfactual scenario, we simulate households' responses to an ad campaign with the "observed" quality in the data and a counterfactual ad campaign of lower quality. We think of a "low" quality ad as one that gives little utility upon consumption, perhaps because it is uninteresting or uninformative. Because we have endogenized ad consumption, our model accounts for the fact that households will be more likely to skip low quality ads and will thus see less ad time. Furthermore, because product consumption is a function of ad time, our model will also predict a decrease in purchase quantity and a reduction in utility from consumption.

We simulate the response to a low quality ad campaign by reducing the baseline utility individuals derive from consuming ads. Specifically, we calculate $\hat{\alpha}_{4i}$ for each household using equation 30 (see Figure 10). The mean and standard deviation of these estimates are 3.311 and 0.355, respectively. Then, to simulate a reduction in quality, we reduce each household's estimate by 2 of these estimated standard deviations.

The plots below show the first-order effects on ad consumption of a reduction in ad quality. Comparing the skip rates reported in Figure 17 to those in Figure 11, we see that skip rates approximately double when the quality of ads is reduced. Figure 18 shows that with this increase in ad-skipping, the mean household reduces their consumption of brand 4 ads by 2%.

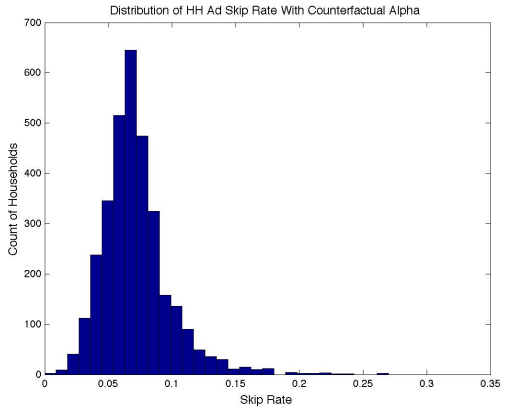


Figure 17: Distribution of Household Ad Skip Rates of Low Quality Brand 4 Ads

This reduction in ad time leads to a reduction in purchase incidence, purchase quantity, and utility derived from consumption. See Figures 19 and 20a. The mean reduction in purchase quantity is 1% and the mean reduction in utility from consumption is 0.04%. Additionally, Figure 20b shows that households derive less utility from the consumption of ads because they consume fewer ads and they derive less utility from the advertisements that they do consume.

At this stage our model does not have an explicit way to predict how changes in ad quality would

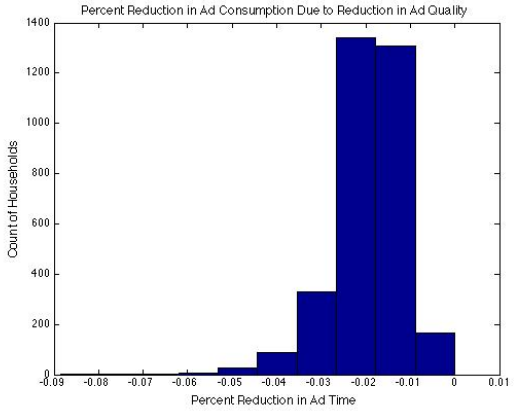


Figure 18: Percent Change in Consumption of Brand 4 Ads Resulting from a Decrease in Ad Quality

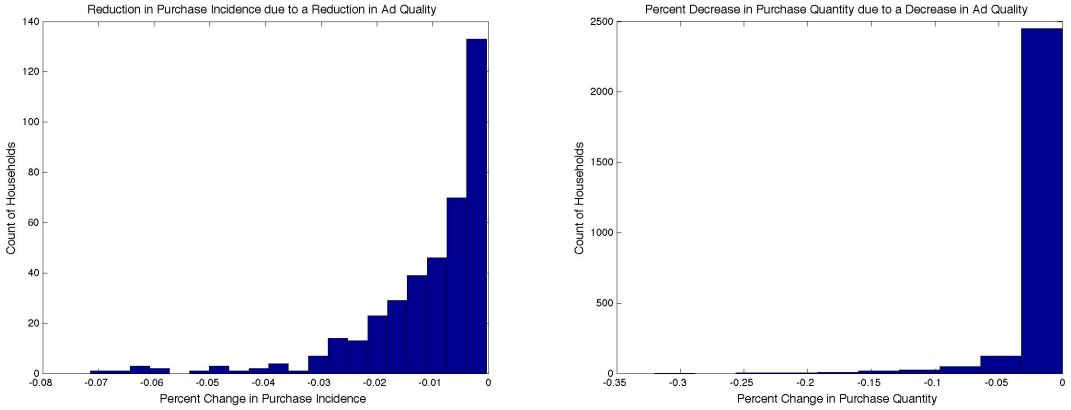


Figure 19: Percent Change in Purchase Incidence and Quantity Due to a Decrease in Ad Quality

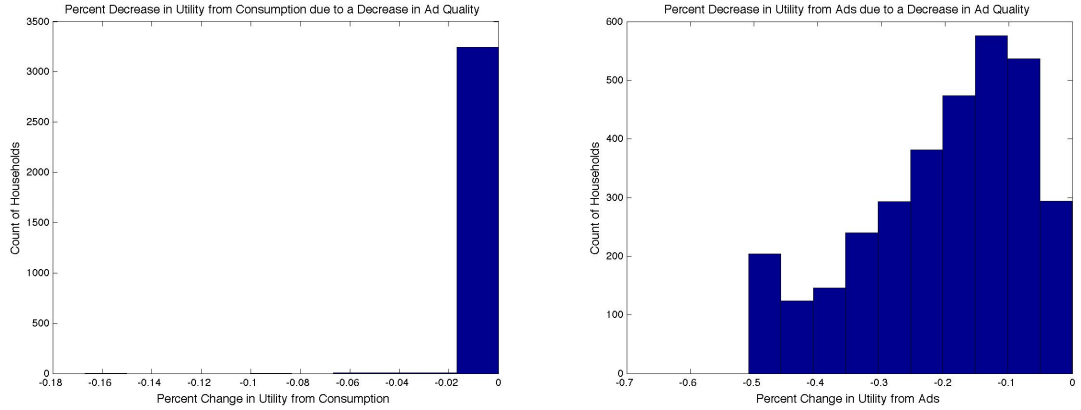


Figure 20: Percent Decrease in Utility from Consumption of Products and Consumption of Ads Resulting from a Decrease in Ad Quality

be manifest in the model's parameters. However, we can imagine two ways in which quality could be directly integrated in future extensions of our model. First, if information on ad content were available, an interesting extension would be to include copy-specific intercepts in the ad consumption equation and project these intercepts on ad characteristics to determine which characteristics are associated with lower skip rates. Then an intercept could be predicted for a new advertisement based on the characteristics of the new ad.[16] Alternatively, firms could pre-test advertisements in focus groups or conjoint tasks and measure observed skip rates of new advertisements. These relative skip rates could be used to calibrate a realistic reduction in the intercept in the ad consumption equation.

We are currently working on other counterfactuals in which we assess the gain from micro-targeting by re-optimizing individual-level advertising and pricing based on the model.

8 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 are allowed to generate complementarities in utility. Using data on detailed product purchase and TV ad-viewing choices by consumers, we document that advertising significantly shifts quantity demanded. We also document that many ads are skipped, and that the skip-rates are explained by the quantity of the advertised product that was purchased recently. 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. The model facilitates assessing these precisely. 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.

We are currently working on various counterfactual simulations to explicate the additional implications of treating advertising in this fashion.

[16] For example, suppose for each ad copy we estimated an intercept $\hat{d}_{jr}$ with asymptotic variance-covariance $\hat{\Delta}$, where r denotes a unique ad copy. We could project these estimates on a matrix of brand and advertising characteristics Z using FGLS s.t. $\hat{\varphi}_{FGLS} = (Z'\hat{\Delta}^{-1}Z)^{-1}Z'\hat{\Delta}^{-1}\hat{d}$.

9 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.
- 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.
- 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.
- Bronnenberg, B., Dube, J-P., Mela, C. (2010). "Do DVRs Influence Consumers' Brand Purchases?" *Journal of Marketing Research*, 47(6), Dec.
- Chintagunta, P., Dube, 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.
- Deng, Y. (2014). "DVR Advertising Targeting," working paper, Fuqua School of Business, Duke University.
- Fox, J. and Lazatti, N. (2014). "Identification of Discrete Choice Models for Bundles and Binary Games," working paper, Univ. of Michigan, Dept. of Economics.
- Gentzkow, M. (2007). "Valuing New Goods in a Model with Complementarity: Online Newspapers," *The American Economic Review*, 97(3), 713-744.
- Greene, M. (2007). "US DVR Forecast, 2007 to 2012," (Dec 28).
- Interactive Advertising Bureau. (2014). "2013 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., Navernioukk, 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, Navernioukk, 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*, April. Available at: `http://static.googleusercontent.com/media/research.google.com/en/us/pubs/archive/35167.pdf`

- Kim, K., Allenby, G.M. and Rossi, P.E. (2002). "Modeling Consumer Demand for Variety," *Marketing Science*, 21(3), 229-250.
- 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.
- Naik, P. A., M. K. Mantrala, A. G. Sawyer. (1998). "Planning Media Schedules in the Presence of Dynamic Advertising Quality," *Marketing Sci*ence, 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.
- 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`
- Story, L. (2006). "A Question of Eye-Balls," *New York Times* (Oct 18), Available at: `http://www.nytimes.com/2006/10/18/business/media/18adco.html`
- 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.
- 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., Navernioukk, I. (2009). "Measuring Advertising Quality on Television Deriving Meaningful Metrics from Audience Retention Data," *Journal of Advertising Research*, 49, 419-428

10 Appendices

A 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 hhid-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 purchased 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.

6. Because barcodes correspond to different package volumes, we create a barcode-store-week level price per unit time series 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 time series by the total volume of that barcode bought at that store over the sample.

8. Finally, we create a brand-hh-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 135075 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.

	(1) Ad Watched 100%	(2) Ad Watched 95%	(3) Ad Watched 75%
$Q_{ijt,14}$	1.8e-05***	1.67e-05**	2.27e-05***
	(6.56e-06)	(6.56e-06)	(7.73e-06)
Time Trend	0.0024***	0.0024***	0.0022***
	(2.8e-05)	(2.9e-05)	(3.23e-05)
Observations	1,436,400	1,436,400	1,436,400
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. $Q_{ijt,14}$ records the cumulative package volume household i purchased of brand j in the 14 days preceding day t.

Table 16: Logit Regression of Ad Watched Dummy on Cumulative Purchase Quantity

B 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 16 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.

π	λ
0.6565	7.9383
(0.0007)	(0.0137)
Brand	φ_{ij}
1	-
	-
2	0.0961
	(0.0466)
3	0.1063
	(0.0503)
4	0.1232
	(0.0564)
5	0.0741
	(0.0384)
6	-
	-
7	0.1552
	(0.0679)
8	0.1496
	(0.0659)
9	0.1106
	(0.0519)
10	0.0956
	(0.0464)
11	0.0892
	(0.0002)

Table 17: Additional Advertising Model Parameter Estimates

C Additional Ad Model Parameter Estimates