Advertising, the Matchmaker

Bharat N. Anand†  Ron Shachar‡
Harvard University  Tel-Aviv University

December, 2001

Abstract

This study models advertising content as a noisy signal on product attributes. Contrary to previous empirical studies that modeled advertising only as part of the consumer’s utility function, we formulate advertising also as an element in her information set. This approach yields the following implications. First, in some cases, exposure to advertising decreases the consumer’s tendency to purchase the promoted product. Second, exposure to advertising improves the matching of consumers and products. These implications enable the researcher to distinguish between the effect of advertising on utility and its effect through the information set using individual-level data on both consumption and advertising exposures. Using a dataset that was designed and created to test this model and its implications, we show that the theory is supported empirically. The structural estimates imply that an exposure to a single advertisement decreases the consumer’s probability of not choosing her best alternative by at least 16%.

†We are deeply grateful to Dmitri Byzalov for excellent research assistance, to Steve Berry, Dennis Carlton, Zvi Eckstein, Michael Keane, Paul Klemperer, Aviv Nevo, Ariel Pakes, Peter Reiss, John Rust, Manuel Trajtenberg, Bruce Weinberg, Ken Wolpin, and to participants in seminars at Econometrics in Tel-Aviv, Haifa University, Harvard University, Hebrew University, MIT, NBER IO Meetings, NBER Productivity Seminar, NYU, Oxford University, University of Pennsylvania, Rochester University, Stanford University, Tel-Aviv University, and Yale University, for comments, and to several colleagues for helpful discussions. Anand is grateful to the Division of Research at Harvard Business School for financial support. Shachar is grateful to the Israel Institute of Business Research at Tel Aviv University Faculty of Management for financial support.
‡Soldiers Field Road, Boston, MA 02163. Phone: (617) 495-5082; Fax: (617) 495-0355; email: banand@hbs.edu.
§Eitan Berglas School of Economics, Tel Aviv University, Tel Aviv, Israel, 69978. Phone: 972-3-640-6311; Fax: 972-3-640-5621; email: rroonn@post.tau.ac.il.
1 Introduction

In this study, we model advertising content as an unbiased noisy signal on product attributes. Contrary to previous empirical studies that modeled advertising only as part of the consumer’s utility function, we formulate advertising also as an element in her information set. This approach yields the following implications. First, in some cases, exposure to advertising decreases the consumer’s tendency to purchase the promoted product. Second, exposure to advertising improves the matching of consumers and products. These implications enable a researcher equipped with individual-level data on both consumption and advertising exposure to distinguish between the two effects of advertising on choices, the first through the utility function and the second through the information set. Using a dataset that was designed and created to test this model and its implications, we show that the theory is supported empirically. The structural estimates imply that an exposure to one advertisement decreases the consumer’s probability of not choosing her best alternative by at least 16%.

Grossman and Shapiro (1984) were the first to identify the role of advertising in matching consumers with products.1 In their setting, advertising conveys full and accurate information about product characteristics. Heterogeneous consumers, who have no source of information other than advertising, seek to purchase the product that best matches their tastes. We follow Grossman and Shapiro in various aspects. In the model, heterogeneous consumers are uncertain about product attributes. They face differentiated products, and advertising conveys information about product attributes. Challenged with the need to take this approach to the data, we modify some of the assumptions and construct a different setting. The setting is of a discrete choice model with consumer preferences formulated in the spirit of Lancaster (1971) and Berry, Levinsohn, and Pakes (1995).2 The modified assumptions are that consumers have additional sources of information other than advertisements, such as word-of-mouth, media coverage, and previous experience. These additional sources are modeled as unbiased noisy product-specific signals. Furthermore, rather than assume that advertising conveys full and accurate information on attributes, it is assumed that advertising is an unbiased noisy product-specific signal as well. The consumer’s prior distribution of a product’s attributes is equal to the distribution of attributes in the market. Moreover, unlike Grossman and Shapiro, we allow advertising to enter the model not only through the information set, but also via a direct effect on the utility function.

This setting yields several testable implications. First, exposure to an advertisement might decrease the consumer’s tendency to purchase the promoted product. This would happen whenever

1While Butters (1977) was the first to formulate the informative role of advertising, he studies homogeneous products and advertisements that convey information about products’ existence, not their attributes.

2Consumer preferences in Grossman and Shapiro (1984) follow the circular city model of Salop (1979).
the match between the promoted product’s attributes and consumers’ tastes is below her match with the average attributes of products in the market. The intuition behind this result is as follows. Without any product-specific signals, the expected utility of the consumer is equal to her match with the average attributes of products in this market. Any product-specific signal that she receives shifts her expected utility (on average) towards her true utility from that product. Advertising is such a product-specific signal. Thus, whenever her match with a product is lower than her average match, her tendency to purchase that product decreases with the number of advertisements that she is exposed to. In Grossman and Shapiro, by contrast, any exposure to an advertisement increases an individual’s tendency to purchase the promoted product. This is because, in their setting, a consumer who is not exposed to advertisements is ignorant about the existence of this firm, and thus her probability of purchasing such a product is zero.

Like Grossman and Shapiro, we also find that informative advertising improves the matching of products and consumers. This is the second testable implication of the model.

The estimation of this model presents two significant challenges. The first is the distinction between the direct effect of advertising on utility and its effect through the information set. The model has several implications that enable the researcher to identify these two effects separately. For example, the first implication above predicts that consumers differ in their response to advertising intensity. Furthermore, these heterogeneous responses are correlated with consumers’ heterogeneous preferences over product attributes. Notice that even without the effect of advertising through the information set, consumer responses to advertising intensity might be heterogeneous. These heterogeneous responses are already accounted for in the standard model. However, the point here is that the augmented model offers another source of heterogeneity. This additional source is what identifies the effect of advertising through the information set.

The large number of unobserved variables in this model presents the second challenge. As researchers, while we observe the number of advertising exposures for each individual, we do not observe the realizations of the signals on product attributes (through advertising, word-of-mouth, media coverage, and previous experience). We also do not observe the individuals’ prior distributions on the attributes of each product. Furthermore, we allow individuals’ preferences for products and firms to differ in unobserved ways. To overcome this problem, we follow Pakes and Pollard (1989) and McFadden’s (1989) approach and use simulation integration. To further reduce the simulation error, we employ importance sampling as described in the Monte Carlo literature (see Rubinstein, 1981). Our importance sampler is similar to the one used in Berry, Levinsohn and Pakes (1995, hereafter BLP). We show that one can reduce the dimensionality of the unobserved by rewriting the expected utilities in a compact way. To speed up the estimation, we employ several simple computational solutions.

In order to estimate this model and test its implications, one needs data on consumption and
exposures to advertising at the individual level. In the last decade, a few research companies (e.g., Nielsen) have created such datasets. The products that they cover are ketchup, yogurt, toothpaste, and coffee. While very useful to examine some theories of advertising, these types of data does not suit our model well for the following reasons. First, these are experience goods. Thus, significant attributes of the products are not well-defined and the match between the consumer’s preferences and the product attributes is largely unobserved. Second, Nielsen data covers television advertising only. Thus, advertisements from other sources (newspapers, radio, etc) are not included in the datasets. Third, it is occasionally difficult to track the different prices that consumers face with respect to each of the products. Prices differ by firms, over time, and across consumers (through coupon schemes).

We created a dataset designed to overcome these difficulties. The products that we chose for this purpose are television shows. Accounting for the cost of leisure in consumption, television shows are clearly one of the most important consumption products. Previous studies have already revealed the key attributes of these products. Thus, one can estimate the value of the match between consumer preferences and product attributes. Furthermore, the price of watching a television show is not product-specific. Finally, almost all the commercials for television shows appear on TV. This enables us to create a comprehensive dataset of exposures to advertisements. Specifically, we obtained Nielsen individual-level panel data on television viewing choices for one week in November, 1995. We created data on show attributes, and recorded all the advertisements for these television shows—also called “previews”, “promotions”, or “tune-ins”—that were aired during that week. Combining our records with the Nielsen panel data and show attribute data gives us the required data to estimate the model.

While the aim of this study is to structurally estimate the parameters of the model, we start our empirical investigation by directly testing the model’s implications. We find that, as expected, individuals’ responses to advertising exposures are a function of their preferences over product attributes. We also test the matching role of informative advertising directly. For this purpose, we construct a crude measure of the individual-product match by interacting the demographic characteristics of individuals with those of show cast members. The match experienced by the individual is then assessed by the match value from her chosen product. Recall that the model predicts that advertising improves the matching of consumers and products. Indeed, we find that the match experienced by the individual is a positive function of the number of advertisements that the individual was exposed to. In these two tests, as in other non-structural examinations that are reported in section 5, we control for the direct effect of advertising on consumers’ utility.

The parameter of interest in the structural estimation is the precision (inverse of the variance) of the noisy advertising signal. If the estimate of this parameter were equal to zero, then the information sets of two individuals who differ only in their exposures to advertising are identical.
In that case, advertising does not have any informational role. The estimate of the precision of advertising signals is positive and statistically different from zero at the 1% significance level. Furthermore, the behavioral impact of advertising signals is substantial as well. In order to evaluate this behavioral impact, we compare the precision of advertisements with those of other sources of information. We find that the precision of two advertisements is equal to the precision of all the other product-specific signals together.

The structural estimates of the match between consumers and products are then used to evaluate the matching role of advertising in several ways. For example, it is shown that an exposure to one advertisement decreases a viewer’s probability of watching a show which yields a below-average match from 0.438 to 0.314. Furthermore, we find that an exposure to a single advertisement decreases the consumer’s probability of not choosing her best alternative by at least 16%.

For each product, a firm obviously intends its advertisements to reach consumers whose response to exposures is the largest. The model presented here identifies rich heterogeneity in consumer responses, and the structural estimates are used to evaluate the targeting strategies of the television networks. These estimates serve to locate the advertising placements that maximize the networks’ profits. It turns out that the strategies (e.g., placing an advertisement for a show in the preceding one) that stand behind these “optimal” locations are very similar to the strategies that characterize the actual locations. Furthermore, in 60 percent of the cases, the actual locations coincide with the “optimal” ones.

The availability of individual-level data on consumption and advertising exposures has generated interesting findings by Erdem and Keane (1996), Ackerberg (2001), and Shum (1999). The modeling approach taken here is similar in one aspect to that in Erdem and Keane (1996)—modeling advertising content as a noisy signal. We differ from their approach in several ways. First, by allowing advertising intensity to enter the utility function, we avoid the danger of misspecifying the model. Second, by focusing on search goods that have observable characteristics, we reveal the matching role of advertising and its potential to deter consumption. This last difference expresses itself in another way. Their identification of information in advertising rests entirely on the structure that the model imposes on the variance of choices by individuals. Using the observable characteristics of products, we have other identifying sources as discussed above.\(^3\)

While Ackerberg (2001) and Shum (1999) also employ individual-level panel data, discrete choice models, and structural estimation, their models of advertising are different from the one presented here. Ackerberg, following Milgrom and Roberts (1986), focuses on an experience good and models advertising intensity as a signal of product quality. Notice that unlike our approach, Ackerberg intentionally focuses on the type of advertising that is referred to in Milgrom and Roberts.

---

\(^3\) An additional difference is that we estimate the precision of the advertising signals. This allows us to directly test the existence of advertising in the information set.
as “having little or no obvious informational content.” Using panel data on choices of different
types of yogurt, and exposures to their television advertising, he shows that consumers who had
experienced the product through past consumption were less responsive to advertisements than were
inexperienced consumers. Although Shum’s (1999) model does not deal with any informational role
of advertising, his findings are somewhat similar to Ackerberg’s. In his model, habit, not experience,
is the source of the different responses by consumers to advertising exposures.\footnote{In an early attempt to use our data (see our 1998 study), we find preliminary evidence for the informational role of advertising. Specifically, we find that consumers’ response to advertising intensity is weaker for well-known products than for newly introduced ones.}

2 The Model

This section introduces the utility function, the information set, and the implications of the model.

We study differentiated products and heterogeneous consumers. Following Lancaster (1971),
we formulate consumer utility over products as a function of individual characteristics and the
attributes of those products. Our discrete choice model has a random utility as in McFadden
(1981). This setting is quite similar to the one presented by Berry, Levinsohn, and Pakes (1995).
Like BLP, products in this model are search goods.\footnote{An individual can know her utility from a search good even without consuming it. See Tirole (1988, page 106).} Individuals are uncertain about product attributes and, like Grossman and Shapiro (1984), advertising is informative.

2.1 The utility

There are $I$ individuals, indexed by $i$, who face $J$ products, indexed by $j$. The no-purchase option
is the $(J + 1)$’th alternative.

The utility from consuming a product is:

$$U_{i,j} = X_j\beta_i + (\eta_j + \varepsilon_{i,j}) + g(N^a_{i,j}) \text{ for } j = 1 \ldots J$$

(1)

The first element of the utility represents the match between the products’ observed attributes, $X_j$, and the preference parameters of the individual, $\beta_i$. The variable $X_j$ is a $K$-dimensional row-vector, and the parameter $\beta_i$ is a column-vector of the same size. The parameter vector $\beta_i$ is a function of observable and unobservable individual characteristics. For example, in the automobile industry, miles per gallon is a product attribute, and income is an individual characteristic. The corresponding $\beta$ parameter is likely to be negative.

The utility is also a function of products’ unobserved attributes. These are represented by the
second element of the utility, $(\eta_j + \varepsilon_{i,j})$. Common effects are captured by the parameter $\eta_j$, while
personal effects are represented by the random variable $\varepsilon_{i,j}$. The parameter $\eta_j$ is often referred to as the “vertical” component of utility, while the element $X_j \beta_i$ is called the “horizontal” component.

The third element of the utility is a positive function, $g(\cdot)$, of the number of advertisements that individual $i$ is exposed to for product $j$, $N_{i,j}^a$. This is the modeling approach adopted by previous empirical studies. Notice that this effect, which was termed “persuasive” by Grossman and Shapiro, was not included in their model. Although we present below a different channel through which advertising affects choices, the $g(\cdot)$ function is included in order to avoid misspecification of the model, and to enable comparison between the standard approach and ours.

The utility from the non-purchase alternative is simply:

$$U_{i,J+1} = \gamma_i + (\eta_{J+1} + \varepsilon_{i,J+1})$$

(2)

where the parameter vector $\gamma_i$ is a function of observable and unobservable individual characteristics. The $(\eta_{J+1} + \varepsilon_{i,J+1})$ terms are analogous to the ones defined above for the first $J$ alternatives.

### 2.2 Information set

Unlike most discrete choice models, we assume that the individual is uncertain about product attributes, $\eta_j$ and $X_j$, and thus about $(\eta_j + X_j \beta_i)$. Since this expression represents the contribution of product attributes to utility, we term it “attribute utility”. We denote this element as $\xi_{i,j}$. Specifically,

$$\xi_{i,j} = \eta_j + X_j \beta_{i}.$$  

(3)

The prior distribution of $\xi_{i,j}$ is:

$$\xi_{i,j} \sim N(\mu_i, \frac{1}{\xi_i})$$

(4)

where, by definition, $\mu_i = E(\eta_j) + E(X_j) \beta_i$. While the individual is uncertain about $\xi_{i,j}$, she knows the expected value and the variance of $\eta_j$ and $X_j$. Indeed, while most consumers are not perfectly familiar with the attributes of each product, it is reasonable to assume that they have a good sense

---

6 Since some of the product attributes are unobserved by the researcher, some components of the match element are unobserved as well. The parameter $\eta_j$ can be thought of as the mean (across individuals) of these unobserved matches and $\varepsilon_{i,j}$ can be thought of as the deviations from that mean.

7 For example, Nevo (2001).

8 Becker and Murphy (1993) provide an interpretation for the effect of advertising on utility. They view advertisements as a good that complements the product being advertised.

9 The normality of the prior distribution results from a normality assumption about $\eta_j$ and $X_j$. 
of the distribution of these attributes in the market. Notice that the expectation $\mu_i$ and the precision $\phi_i$ differ across individuals because the taste parameter $\beta_i$ is individual-specific.

The individual receives product-specific signals on product attributes from various sources such as word-of-mouth, previous experience with the product, media coverage, and advertising. In order to focus on the informational role of advertising, we separate the advertising signals from the miscellaneous ones.

The individual receives $N_{i,j}^m$ miscellaneous product-specific signals. These signals are i.i.d. Specifically, each signal is distributed as:

$$\tilde{S}_{i,j,n}^m = \xi_{i,j} + \tilde{\omega}_{i,j,n}^m \text{ for any } 1 \leq n \leq N_{i,j}^m \text{ where } \tilde{\omega}_{i,j,n}^m \sim N(0, \frac{1}{\phi_m}).$$

(5)

We assume that these signals are noisy ($\frac{1}{\phi_m} > 0$) and unbiased. The noisiness can result from various sources. For example, even previous experience is not a precise signal because of limited memory and other human information-processing mechanisms.

The content of each advertisement serves the individual as a signal on product attributes. These signals are i.i.d. Specifically, each signal is distributed as:

$$\tilde{S}_{i,j,n}^a = \xi_{i,j} + \tilde{\omega}_{i,j,n}^a \text{ for any } 1 \leq n \leq N_{i,j}^m \text{ where } \tilde{\omega}_{i,j,n}^a \sim N(0, \frac{1}{\phi_a}).$$

(6)

We assume that the signals are noisy, that is $\frac{1}{\phi_a} > 0$. The noisiness of advertising is well-documented in Jacoby and Hoyer (1982). Using a survey of 2,700 consumers about the content of 60 thirty-second televised communications (including advertisements), they find that 29% of these were miscomprehended by consumers. We assume that the signals are independent for two reasons: (1) firms occasionally use different advertisements for the same product; (2) different exposures to the same advertisement can lead to different impressions. The independence assumption does not affect our qualitative results.

The effect of advertisements through the information set is captured by $\tilde{S}_i$. If $\tilde{S}_i = 0$, then

---

10 For example, while it is hard to stay informed about the attributes of each automobile, most consumers know the distribution of miles per gallon and car size in this industry. In some cases, their knowledge is likely to be more extensive. For example, Japanese automakers are known to produce gasoline-efficient cars whereas Swedish producers are perceived to focus on safety. In the empirical model, we allow such firms’ profiles to enter the prior distribution.

11 Consumer learning through the miscellaneous sources has been the focus of various studies (Crawford and Shum 2000 studied dynamic learning through past experience; Muller, Mahajan and Wind 2000 presents the network effects of word-of-mouth; and Bond and Kirshenbaum 1998 described the effect of media coverage). Our focus is on advertising signals and thus these processes are degenerate in this model.

12 They find similar results in their 1989 study, which uses a survey of 1,250 consumers who were exposed to print ads.

13 The unbiasedness assumption rests on truth-in-advertising regulations. Furthermore, if a firm has an incentive to bias the content of its advertisements, a rational consumer would account for it, and is likely to neutralize the bias. We do not model this game in order to keep the model focused on its key elements.
advertisements are too noisy to convey any information about product attributes. In other words, when $\varsigma^a = 0$ the information sets of two individuals who differ only in $N^a$ are the same. On the other hand, when $\varsigma^a > 0$, the information sets of such consumers differ. Thus, $\varsigma^a$ is the parameter of interest in the empirical study.

### 2.3 Expected utility

Since the only element in the utility that the individual is uncertain about is her “attribute utility”, $\xi_{i,j}$, we start by calculating her expected attribute-utility.

Individual $i$ updates her prior using the product-specific signals to form her expected attribute-utility, $\mu_{i,j}^P$: \(^{14}\)

$$
\mu_{i,j}^P = \frac{1}{\varsigma_{i,j}} \left[ \varsigma_i^\mu \mu_i + \varsigma^a \sum_{n=1}^{N_{i,j}^a} S_{i,j,n}^a + \varsigma^m \sum_{n=1}^{N_{i,j}^m} S_{i,j,n}^m \right] 
$$

(7)

where $\varsigma_{i,j}^P = \varsigma_i^\mu + N_{i,j}^a \varsigma^a + N_{i,j}^m \varsigma^m$, and $S_{i,j,n}^a$ and $S_{i,j,n}^m$ are the realizations of the signals. Notice that $\frac{1}{\varsigma_{i,j}}$ is the variance of her posterior distribution.

Since $S_{i,j,n}^a = \xi_{i,j} + \omega_{i,j,n}^a$ where $\omega_{i,j,n}^a$ is the realization of $\omega_{i,j,n}^a$, and $S_{i,j,n}^m = \xi_{i,j} + \omega_{i,j,n}^m$ where $\omega_{i,j,n}^m$ is the realization of $\omega_{i,j,n}^m$, we can re-write equation (7) as:

$$
\mu_{i,j}^P = \left[ \theta_{i,j} \mu_i + (1 - \theta_{i,j}) \xi_{i,j} \right] + \omega_{i,j} 
$$

(8)

where $\theta_{i,j} = \frac{\varsigma_i^\mu}{\varsigma_{i,j}}$, and $\omega_{i,j} \equiv \frac{1}{\varsigma_{i,j}} \left( \varsigma^a \sum_{n=1}^{N_{i,j}^a} \omega_{i,j,n}^a + \varsigma^m \sum_{n=1}^{N_{i,j}^m} \omega_{i,j,n}^m \right)$.

> From equation (8), one can see that with a finite number of product-specific signals, $\mu_{i,j}^P \neq \xi_{i,j}$. In other words, advertising does not resolve all the uncertainty that the individual faces. Notice that this is one of the differences between this model and Grossman and Shapiro (1984).

The individual is not fully informed in this case because $\theta_{i,j} > 0$ and $\omega_{i,j}$ is not equal to 0.\(^{15}\)

Recall that, by definition, $\mu_{i,j}$ is equal to $E(\eta_j) + E(X_j) \beta_i$. In other words, $\mu_{i,j}$ can be thought of as the expected utility from a hypothetical product whose attributes are equal to the mean of the distribution in the market. The reliance of the individual on $\mu_{i,j}$, which is implied by $\theta_{i,j}$, indicates that she is not fully informed about the attributes of the specific product. Thus, one can consider $\theta_{i,j}$ as a measure of how ill-informed the individual is about product attributes.

The weight $\theta_{i,j}$ is a negative function of $N_{i,j}^a$ and $\varsigma^a$. Since advertising is informative, an increase in the number or precision of advertisements would increase the informedness (thus, de-\(^{14}\)See DeGroot [1989].

\(^{15}\)The probability that $\omega_{i,j} = 0$ is equal to 0.
creasing \( \theta_{i,j} \) of the individual. The effect of \( N_{i,j}^a \) on \( \theta_{i,j} \) is a function of \( \zeta^m, \zeta^u, \) and \( N^m \). In other words, the informational effect of advertising is smaller in the following cases: (1) the variety of attributes in the market is smaller (\( \zeta^u \) is larger), (2) the other product-specific signals provide more information (\( N_{i,j}^m \) is larger).

The expected attribute-utility is a negative function of \( N_{i,j}^a \) when \( \zeta_i > \zeta_{i,j} \) and a positive function otherwise. To see this, recall that \( \theta_{i,j} \) is a negative function of \( N_{i,j}^a \). Thus, an increase in the number of advertisements decreases the weight on \( \mu_i \) and increases the weight on the actual attribute-utility. Whenever \( \mu_i > \zeta_{i,j} \), such an increase in the number of advertisements leads to a decrease in the expected attribute-utility. Later, we build on this result and show that informative advertising can deter consumption. Furthermore, this result reveals the matchmaking role of advertising, in other words the role of advertising in improving the match between individuals and products.

The expected utility of the individual is a function of \( \mu_{i,j}^p \), the persuasive effect of advertising, and \( \varepsilon_{i,j} \). Specifically,

\[
E[U_{i,j} | IS_{i,j}] = (\mu_{i,j}^p + \varepsilon_{i,j}) + g(N_{i,j}^a) \tag{9}
\]

where \( IS_{i,j} \) is the information set of individual \( i \) on product \( j \), and \( IS_{i,j} = \{ \mu_i, \{ S_{i,j,n}^a \}, \{ S_{i,j,n}^m \}, \zeta^u, \zeta^m, \zeta^a \} \).

It is easy to show that the probability that individual \( i \) will choose alternative \( j \) is a positive function of her expected utility from that alternative.

### 2.4 Implications

This model has several testable implications. In order to derive these, we define the choice probabilities from the standpoint of a researcher.

While the individual observes the realizations of the signals, but not \( \zeta_{i,j} \), the researcher does not observe the signals but has \( X_j \) and estimates of \( \eta_j \) and \( \beta_i \), and thus an estimate of \( \zeta_{i,j} \). Since the signals are unobserved, \( \omega_{i,j} \) is a random variable from the researcher’s point of view. It is distributed normally with mean 0 and standard deviation \( \sigma_{i,j}^\omega = \sqrt{\zeta_u N_{i,j}^a + \zeta^m N_{i,j}^m / \zeta_{i,j}^2} \). In order to write the expected utility from the researcher’s point of view, we replace \( \omega_{i,j} \) by \( \sigma_{i,j}^\omega z_{i,j} \) where \( z_{i,j} \) is a standard normal random variable. The expected utility is then:

\[
u_{i,j} = [\theta_{i,j} \mu_i + (1 - \theta_{i,j}) \zeta_{i,j}] + \sigma_{i,j}^\omega z_{i,j} + g(N_{i,j}^a) + \varepsilon_{i,j} \text{ for } j = 1, \ldots, J \tag{10}\]

The standard deviation \( \sigma_{i,j}^\omega \) in equation (10) is another measure (in addition to \( \theta_{i,j} \)) of how ill-informed the individual is. When \( \sigma_{i,j}^\omega = \theta_{i,j} = 0 \), the expected utility is exactly equal to the utility. This would happen, for example, if the miscellaneous signals are not noisy (\( \frac{1}{\zeta^m} = 0 \)). In this case,
the individual is fully informed.

The following notations and assumptions simplify the subsequent presentation. We simplify the model without loss of generality by replacing $N_{i,j}^m \gamma^m$ with $\zeta_{i,j}$. Notice that $N_{i,j}^m$ and $\gamma^m$ always appear in the model as $N_{i,j}^m \gamma^m$. Denote as $\zeta_i^m$ the $J$-element vector whose $j$'th component is $\zeta_{i,j}$. Accordingly define $\varepsilon_i$ and $z_i$.

In addition to $z_i$, the researcher does not observe $\bar{\beta}_i$, $\gamma_i$, and $\zeta_i^m$. Let $\vec{\omega}_i$ be the set of all parameters that are common across individuals, and let $A_{i,j}(W_{i,j}, \Omega) = \{ (\varepsilon_i, v_i, z_i) \mid u_{i,j} \geq u_{i,r}, \text{ for } r = 1, ..., J + 1 \}$. (12)

That is, $A_{i,j}$ is the set of values for the variables and parameters that are unobserved by the researcher that induces the choice of product $j$. Then, the probability that individual $i$ chooses alternative $j$, given the parameters, is:

$$p_{i,j} = \int_{A_{i,j}} dP^*(\varepsilon, v, z)$$

where $P^*(\cdot)$ denotes the distribution functions.

Advertising affects $p_{i,j}$ through the information set and also via the utility. In order to identify the implications of each of these channels, we start the analysis by assuming that $g'(\cdot) = 0$. This allows us to focus on the consequences of informative advertising.

Implication 1 Assuming that $g'(\cdot) = 0$, $p_{i,j}$ is decreasing in $N_{i,j}^a$ if $\xi_{i,j} < \mu_i$, and increasing otherwise.

As mentioned above, the expected attribute-utility is a negative function of $N_{i,j}^a$ when $\xi_{i,j} < \mu_i$. Since the set $A_{i,j}$ increases when the expected attribute-utility increases, we get this implication.

The intuition behind this result is simple. Whenever the match between a consumer and a product is low, any product-specific information will decrease the consumer’s tendency to buy the
product. Advertising provides such information.\textsuperscript{16}

Since the sign of $\xi_{i,j} - \mu_i$ depends on $\beta_i$, the implication above means that the informative effect of advertising depends on consumer taste parameters. In contrast, the persuasive effect, through $g(\cdot)$, does not. This difference between the effect of advertising through the information set and via the utility enables a researcher to empirically distinguish between the effects. In the data, these two effects would exist together, and thus one expects to find that responses to advertising are positive on average, heterogeneous across consumers, and that this heterogeneity depends on $\beta_i$.

Notice that in Grossman and Shapiro (1984), any exposure to an advertisement increases an individual’s tendency to buy the promoted product. The reason is that without any advertisements, the consumer is ignorant about the existence of this firm, and thus her probability of buying such a product is zero.

Like Grossman and Shapiro (1984), we also find that informative advertising improves the match between products and consumers. This is the second testable implication of the model.

**Implication 2** Assuming that $g'(\cdot) = 0$, $\sum_{j=1}^{J+1} U_{i,j} p_{i,j}$ increases with $N_{i,j}^a$ for any $j$.

This means that an advertisement about any product improves the matching process. Again, the intuition is straightforward. By reducing consumers’ tendencies to purchase products which do not fit their preferences well, and increasing their tendency to buy those that do, advertising increases $\sum_{j=1}^{J+1} U_{i,j} p_{i,j}$.

There are other testable implications of the model. First, the effect of $E(X_j)$ on $p_{i,j}$ is decreasing in $N_{i,j}^a$. Recall that the reliance of an individual on market information, $E(X_j)$, results from her uncertainty on a product’s attributes. Since advertising partially resolves this uncertainty, her dependence on $E(X_j)$ diminishes. Second, the conditional correlation between choices and product attributes depends on $N_{i,j}^a$. This correlation depends negatively on the variance of the variables which are unobserved by the researcher. In the model, this variance depends on $\sigma_{i,j}^\omega$. Since $\sigma_{i,j}^\omega$ is a function of $N_{i,j}^a$, we get this implication.

Extending this one-period model to a multiperiod setting, as we do in section 4, reveals two additional implications which are discussed then.

### 3 Data

The empirical application of this model comes from the television industry. The data include product attributes, individual characteristics, individual (television viewing) choices, and individual-

\textsuperscript{16}The product-specific signals are noisy. Thus, a signal might decrease a consumer’s tendency to buy a product even when $\xi_{i,j} > \mu_i$. However, since the expected value of $\omega_{i,j,n}$ is zero, this idiosyncratic effect cancels out in $p_{i,j}$.  

12
level exposures to advertisements (promoting television shows). The data on individual characteristics and choices for the week that starts on Monday, November 6, 1995, were obtained from A.C. Nielsen, and the rest of the data were designed and created for the purpose of this study.

Previous individual-level studies of advertising relied on a dataset of consumption and exposures to advertising that was put together by Nielsen. This dataset consists of four product categories: yogurt, ketchup, toothpaste, and coffee. We start by describing the shortcomings of this dataset, which led us to create the new dataset.

3.1 Suitability of the data

The empirical task demands that the data satisfies the following requirements: (1) products are differentiated, (2) consumers are heterogenous, (3) consumers are uncertain about product attributes, and (4) the researcher observes some of the product attributes. Previous studies show that data on television viewing choices satisfy these requirements. The most important deviation of the data on yogurt, ketchup, toothpaste, and coffee from these requirements is the lack of observable product attributes over which consumers’ tastes vary.

Another disadvantage of the exposure data created by Nielsen is that it does not include exposures to advertisements that appear in newspapers and radio. Indeed, our data would appear to suffer from the same problem—we only observe advertisements that appear on TV. However, in our case the problem is not severe since almost all advertisements for television shows appear on TV. This is not the case for the other products mentioned earlier.

The data put together by Nielsen raises another difficulty for researchers—the use of coupons. Specifically, the decision to use a coupon is endogenous and the availability of a coupon is unobserved. This problem is avoided in our data, since the monetary cost of viewing a network television show is zero, and the non-monetary cost is the same (for each individual) across shows in any period.

The last advantage of data on TV is that viewing television shows is an important consumption activity. On average, an American watches television for four hours per day. Accounting for the opportunity cost of leisure, spending on television consumption is high.

---

17 Rust and Alpert (1984) and Shachar and Emerson (2000) identify product attributes and demonstrate that consumers’ tastes for these attributes vary in the population. Anand and Shachar (2001a) show that viewers are uncertain about product attributes. While basic attributes, such as whether a television show is a comedy or not, may be easily discernible from the television schedule that appear in daily newspapers, other attributes, such as the level of romance in a particular episode, are not available. Furthermore, the focus of a show frequently shifts from one episode to another. For example, one episode might focus on a female character and her personal dilemmas, while the next is centered around her male spouse.

18 Anderson and Coate (2000) cite data from the Television Advertising Bureau that the average adult man in the U.S. spent 4 hours and 2 minutes watching television per day, and the average woman spent 4 hours and 40 minutes per day.
3.2 The data sets

The datasets are presented in the following order: product attributes, consumer characteristics, consumption choices, and exposures to advertisements.

3.2.1 Product (Show) Characteristics

We coded the show attributes for the 64 shows in the relevant week based on prior knowledge, publications about the shows, and viewing each one of them. Following previous studies, we categorize shows based on their genre and their cast demographics. Rust and Alpert (1984) present five show categories—for example, comedies and action dramas—and show that viewers differ in their preferences over these categories. We use the following categories: situational comedies, also called “sitcoms” (31 shows fall into this category), action dramas (10 shows), and romantic dramas (7 shows). The base group includes news magazines and sports events (16 shows), which was found by previous studies to be similar.19

Shows were also characterized by their cast demographics. Shachar and Emerson (2000) demonstrate that the demographic match between an individual and a show’s cast plays an important role in determining viewing choices. For example, younger viewers tend to watch shows with a young cast, while older viewers prefer an older cast. We use the following categories: Generation-X, if the main characters in a show are older than 18 and younger than 34 (21 shows fall into this category); Baby Boomer, if the main show characters are older than 35 and younger than 50 (12 shows); Family, if the show is centered around a family (11 shows); African-American (7 shows); Female (15 shows); and Male (22 shows).

3.2.2 Consumer Characteristics and Choices (The Nielsen Data)

We obtained data on individuals’ viewing choices and characteristics from Nielsen Media Research. Nielsen maintains a sample of over 5,000 households nationwide.20 Nielsen installs a People Meter (NPM) for each television set in the household. The NPM records the channel being watched on each television set. A special remote-control records the individuals watching each TV. Thus, the viewing choices are individual-specific. While criticized occasionally by the networks, Nielsen data still provide the standard measure of ratings for both network executives and advertising agencies.

Although the NPM is calibrated for measurements each minute, the data available to us provide quarter-hour viewing decisions, measured as the channel being watched at the midpoint of

---

19See Goettler and Shachar (2001).
20Using 1990 Census data, the sample is designed to reflect the demographic composition of viewers nationwide. The sample is revised regularly, ensuring, in particular, that no single household remains in the sample for more than two years.
each quarter-hour block. Our data consists of viewing choices for the four major networks, ABC, CBS, NBC, and Fox.

We focus on viewing choices for network television during prime time, 8:00 to 11:00 PM, using Nielsen data from the week starting Monday, November 6, 1995. Thus, we observe viewers’ choices in 60 time slots. Figure 1 provides the prime-time schedule for the four networks over this week. This study confines itself to East coast viewers, to avoid problems arising from ABC’s Monday night programming.21 Finally, viewers who never watched television during weeknight prime time and those younger than six years of age are eliminated from the sample. From this group, we randomly selected individuals with a probability of 50 percent. This gives us a final sample of 1675 individuals. On average, at any point in time, only 25 percent of the individuals in the sample watch network television.

In addition to viewer choices, Nielsen also reports their personal characteristics. Our data includes the age and the gender of each individual, and the income, education, cable subscription and county size for each household. Table 1 defines the variables created based on this information, and their summary statistics.

3.2.3 Data on exposures to advertising

We taped all the shows for the four networks during the week that starts on November 6, 1995. We then coded the appearance of each advertisement for the television shows. For example, on Monday at 9:10 PM, there was an advertisement for the ABC newsmagazine 20/20 (this show aired on Friday at 10:00 PM). This information was matched with the Nielsen viewing data to determine an individual’s exposure to advertisements. For example, an individual who watched ABC on Monday at 9:10 PM was exposed to the advertisement mentioned above. Summing over all time slots, we get the number of exposures of individual \( i \) with respect to each show in the week. In 1995, these advertisements, which are also referred to as “promos”, usually included the broadcast time of the show, and clips from the actual episode.

Since our Nielsen viewing data starts on Monday we cannot determine the exposure to advertisements that were aired before that day. This means that our data miss some exposures for shows. This problem is likely to affect the exposure variable for shows which were broadcast on Monday and Tuesday, and less likely to influence those which aired on Wednesday through Friday. Thus, in the non-structural tests, we use only the data for Wednesday through Friday, and in the structural estimation, we allow the advertising parameters to differ across these two parts of the week.

\footnote{ABC features Monday Night Football, broadcast live across the country; depending on local starting and ending times of the football game, ABC affiliates across the country fill their Monday night schedule with a variety of other shows. Adjusting for these programming differences by region would unnecessarily complicate this study.}
For the Wednesday through Friday shows, the mean number of advertisements aired per show is 4.9, and the median is 4.  On average, an individual is exposed to 0.37 advertisements for each show on Wednesday through Friday. Since some people do not watch any show on Monday and Tuesday, a more meaningful measure of exposures to advertisements is given by conditioning on watching television in more than two time slots during these two days. In this case, the average exposure is 0.56.

4 Preliminary evidence

In order to separate the informative effect of advertising from its persuasive effect, we have drawn the model’s implications under the assumption that $g'(\cdot) = 0$. In the empirical examination, we obviously cannot make this assumption. In the structural estimation, we simultaneously estimate the parameters of the $g(\cdot)$ function and $\zeta^a$. Here, we offer a non-structural approach to distinguish between the two effects.

Recall that the implications predict the change in consumer behavior from an increase in the number of exposures. The difficulty arises from the fact that such an increase alters not only the information set of consumers, but the persuasive component of advertising as well. A way to resolve this problem is by equalizing the persuasive effect over all the alternatives. To do that, we focus on a subsample of observations in which each individual was exposed to the same number of advertisements for all the competing shows in a time slot. That is, we compare viewers who were exposed to zero advertisements for each of the competing shows in a specific time slot, with people who were exposed to one advertisement for each of those shows, etc.

Advertising may deter consumption  Here, we provide some preliminary evidence that consumer responses to advertising are a function of their match with a product. Specifically, an increase in the number of exposures lowers the viewing probability for consumers who have a poor match with a product, while raising the viewing probability for those who have a good match with it.

Our non-structural approach to control for the persuasive effect requires that we focus on time slots promoted heavily by the networks. If not, our sub-sample of consumers who are exposed to even one advertisement for each alternative might be too small. It turns out that, for some idiosyncratic reason, the three shows with the highest number of advertisements were aired in the same time slot, Thursday at 10:00 P.M.  The second challenge that the non-structural approach presents is the assessment of the match between consumers and products. It turns out that one

---

22 Promos represent about one of every six minutes of advertising time on the broadcast networks (see Shachar and Anand 1998). Thus, ad-sales ratios are about 16% for the networks.

23 ABC placed 9 ads during the week for its show Murder One, CBS aired 10 ads for its show 48 hours, and NBC placed 8 ads for E.R. Notice that FOX does not offer national programming after 10:00 PM.
of the shows in this time slot is a newsmagazine (48 Hours on CBS). News-magazine is a very clear category. As a result, it is relatively easy to identify individuals who have a good and a bad match with this show based on their viewing choices during the rest of the week. We split the population into two groups of viewers: those who have seen more newsmagazines during the rest of the week than the average viewer, and those who have seen less. For each of these groups, we then compare the viewing probability of those who have been exposed to either 0 or 1 advertisement for each network, with those who have seen 2 or more advertisements for each network. Table 2 demonstrates that when the number of exposures increases, the tendency to watch 48 Hours falls for viewers who dislike newsmagazines, and increases for those who like this category. Specifically, for viewers who dislike newsmagazines, the probability of watching 48 Hours decreases from 5.9% for those who were exposed to 0 or 1 promos, to 4.3% for those who were exposed to 2 or more promos (there are 2164 and 36 individuals in the two categories, respectively). On the other hand, for viewers who like newsmagazines the propensity to watch 48 Hours increases from 16.7% to 28.1% (there are 96 and 68 individuals in the two categories, respectively).

This simple table hints at two additional behavioral features. The first relates to the information set of consumers. The top row of table 2 shows that the tendency to watch the show among individuals who were exposed to one advertisement or less, is 5.9% for those who dislike newsmagazines and 16.7% for those who like this type of show. This suggests that advertising is not the only product-specific signal. In other words, consumers are somewhat informed about the attributes of each product even without any exposures to advertisements. The second feature concerns firms’ strategic behavior. The number of observations in each cell reveals that 41% of individuals who like newsmagazines were exposed to more than one advertisement for the show, compared with only 1.6% among those who dislike newsmagazines. In other words, consumers who like this type of show are more likely to be exposed to advertisements promoting the show. We study such strategic considerations (in targeting advertisements) in section 7.

While table 2 is rich in behavioral features, its statistical power is not. For example, the decrease in the tendency to watch 48 Hours among those who dislike newsmagazines (from 5.9% to 4.1%) is not statistically significant even at the 10% level. Table 3 builds on the logic of table 2. While still a descriptive table, its statistical power is stronger. There are three newsmagazines in the schedule on Thursday and Friday (48 Hours, Dateline on NBC at Friday 9:00 PM, and 20/20 on ABC at Friday 10:00 PM). Let k index these three time slots. Table 3 pools all the observations of table 2 with those of the other two shows. The dependent variable, f_i,k, in this regression is a transformation of the percentage of time spent by individual i watching the newsmagazine alter-

24 As discussed later, there is state dependence in viewing choices. This introduces an additional challenge in the non-structural approach. To overcome this problem, we focus only on individuals who watched TV on Thursday at 9:45 PM.
native in period $k$, denoted as $Watch_{i,k}$. Specifically, $f_\_Watch_{i,k} = \ln(\frac{Watch_{i,k}}{1 - Watch_{i,k}})$. Among the independent variables, there are five that serve as controls and three that examine the consumption-deterring hypothesis. These three are $Ad_{i,k}$, $NewsMatch_{i,k}$, and $Information_{i,k}$. The first, $Ad_{i,k}$, is a binary variable that is equal to one if the number of advertising exposures of individual $i$ to each one of the shows in time slot $k$ is 2 or higher, and is equal to zero otherwise. $NewsMatch_{i,k}$ is an individual-specific taste measure for newsmagazines, constructed as the number of timeslots that the individual watched a newsmagazine divided by the number of timeslots during which a newsmagazine aired. $NewsMatch_{i,k}$ excludes the newsmagazine in timeslot $k$ from the numerator and denominator. Finally, $Information_{i,k} = Ad_{i,k}(1 - NewsMatch_{i,k})$. We expect the effect of the first two variables to be positive, and of the third to be negative. The effect of $Ad_{i,k}$ and $Information_{i,k}$ would imply that the response of consumers to advertisements depends on their tastes. Indeed, we find that the data support this hypothesis.

Advertising improves matching Here, we provide some initial evidence that exposures to advertising improves the matching between consumers and products. Specifically, we examine the relationship between $\sum_{j=1}^{J+1} U_{i,j}p_{i,j}$ and exposures to advertising. Notice that Implication 2 suggests that this relationship should be positive.

Since we have not yet estimated the model’s parameters, we do not have an estimate of $U_{i,j}$. Therefore, we construct a variable $Match_{i,j}$ which represents the demographic match between viewers and shows. This variable is based on three demographic characteristics: age, gender, and family status. It counts the number of characteristics that are identical for both the show and the individual. For example, for a Generation-X single female viewer and a Generation-X show with a single, male cast, $Match_{i,j} = 2$. The test, then, is to examine the relationship between $Match_{i,j}^C \equiv \left( \sum_{j=1}^{J+1} I\{C_i = j\} \cdot Match_{i,j} \right)$ and the number of advertisements that the individual is exposed to. Notice that we have replaced $p_{i,j}$ with $I\{C_i = j\}$ because we do not have an estimate of the probability yet. Thus, one can think of $Match_{i,j}^C$ as the match with the alternative that is chosen by the individual.

\footnote{Whenever $Watch_{i,k} = 0$, $\ln(\frac{Watch_{i,k}}{1 - Watch_{i,k}})$ is set to be equal to $\ln(10^{-6})$. Similarly, when $Watch_{i,k} = 1$, $\ln(\frac{Watch_{i,k}}{1 - Watch_{i,k}})$ is set equal to $\ln(10^6)$.}

\footnote{The other variables account for individual tendencies to watch TV, state dependence, and show fixed effects. Specifically, $All_{i,k}$ is the average number of timeslots that individual $i$ watched TV during all the days of the week excluding the day of timeslot $k$. $Same_{i,k}$ is similar to $All_{i,k}$, but is based only on the same timeslot as $k$. For example, for 48 Hours, only observations from 10:00 PM are included. $LeadIn_{i,k}$ is a binary variable that is equal to 1 if individual $i$ watched the same network in the timeslot just before the show started. The fixed effects are captured by two dummy variables for the shows 20/20 and Dateline.}

\footnote{The age match is slightly different. There are four age groups: teens, generation X, baby boomers and older. The $Match_{i,j}$ variable gets a value of one when the age group of the individual and the show are the same. Otherwise, the index is equal to one minus one half the number of age groups that separate the age group of the individual and the show’s cast. Thus, for example, the match value of a teen watching a generation X show is 0.5.}
Once again, we conduct this test only for individuals who have been exposed to the same number of advertisements for all the networks, and denote this number as $N^a_i$. Table 4a demonstrates that $\text{Match}^C_i$ indeed increases in $N^a_i$. However, notice that the number of observations with positive $N^a_i$ is small. In order to have a more powerful test, we restrict the analysis to considering pairs of shows from the three leading networks. Moreover, since (as we show later) lagged choices affect utility, we separately examine the cases that correspond to different lagged choices (i.e., not watching the networks, watching one, or the other). The results are presented in tables 4b-4d. In most of the cases, the implication is supported by the data. In addition, the support is the strongest (and most significant) for the cases with the largest number of observations with positive $N^a_i$. Consider, for example, the case comparing CBS and NBC shows for individuals who watched NBC in the previous time slot (table 4d). We find that the match increases from 0.497 for individuals who were not exposed to any advertisements for these shows, to 0.546 for those who were exposed to exactly one advertisement, and 0.593 for those who were exposed to at least two advertisements. These differences are significant at approximately the 1% level.

Advertising reduces regret This subsection exploits a special feature of the data that enables us to examine the hypothesis of informative advertising in an additional non-structural way. We exploit the fact that each television show spans multiple 15-minute time slots. Thus, we observe whether viewers who watched the first 15 minutes of a show stick with it or switch away. An individual may have various reasons to switch away from a show after starting to watch it. One of these is that she finds out that the show does not fit her preferences well. Our model suggests that viewers who have been exposed to many advertisements are more informed, and thus less likely to be disappointed in this way. Thus, we expect that individual $i$’s tendency to switch away from a show decreases in $N^a_i$. Indeed, the correlation between the switching decisions and $N^a_i$ is equal to -0.45. This calculation considers only individuals who have seen the same number of advertisements for all the networks and shows on Wednesday through Friday.

This result might be a statistical artifact—we might have omitted a variable that is correlated with both the exposure to advertising and the switching decisions. The personal taste for television might be such a variable. A person who dislikes watching television is likely to be (a) exposed to few advertisements, and (b) switch away (to the outside alternative) more often. Table 5 includes this control variable. The dependent variable in this probit model is equal to one for individuals who switched away from a show after viewing its first 15 minutes, and equal to zero otherwise. The variable $All_i$ measures the fraction of time that individual $i$ watched TV, excluding the night of the show in question. Controlling for viewing time, the data still support the hypothesis. Specifically, we find that the switching rate drops by 2.5% with exposure to an additional advertisement. This
5 Estimation and Identification Issues

This section consists of five subsections. The first, (5.1), extends the model presented in section 2 to account for the multiperiod nature of the data. The specific functional forms of the utility, and the density functions of the unobserved variables, are presented in 5.2. The next subsection (5.3) constructs the likelihood function, and our simulation approach is described in 5.4. The final subsection (5.5) discusses the identification of the model’s parameters.

5.1 The Multiperiod Model

In any given week, television networks offer multiple shows, with each show being offered only once. Recall that in the simple model, each firm offers one product and the consumer chooses among the alternatives once. Like the simple model, the television viewer still faces $J$ products at any point in time. The differences are that each firm (television network) has multiple products and each consumer chooses among the alternatives several times. This subsection extends the model presented in section 2 to account for these features of the data.

5.1.1 Utility

Previous studies of television viewing choices find strong evidence of state dependence. Indeed, our data reveals that on average 65 percent of viewers who were watching a show on network $j$ watch the next show on the same network. State dependence is obviously not the only explanation of this finding. The networks tend to schedule similar shows in sequential time slots. This strategy might also lead to the high persistence rate in choices. However, it turns out that controlling for observed and unobserved show attributes does not eliminate the support for state dependence (Goettler and Shachar, 2001). While there is no good behavioral explanation for state dependence in television viewing, we allow the utility to include a state dependence variable in order to avoid misspecification of the model. Furthermore, in order to avoid biased estimates, we include a network-individual unobserved match variable ($\alpha_{i,j}$) in the utility. The utility function is now time-specific. Specifically,

---

28 Moreover, the negative effect of $N_{i,t}$ on the switching probability is stable across various model specifications, including: (a) when we do not restrict our attention to individuals who have seen the same number of ads for all the networks, and (b) when we focus on the three leading networks only.
29 See Rust and Alpert (1984).
30 This strategy, termed “homogeneity”, is followed by the networks mostly from 8:00 to 10:00 PM. In other words, in most cases, the shows that start at 10:00 PM are dissimilar to those that preceded them. Furthermore, even between 8:00 and 10:00 PM, one can find deviations from this strategy.
the utility of individual $i$ from alternative $j$ in period $t$ (where any combination of $j$ and $t$ defines a show) is:

$$U_{i,j,t} = X_{j,t} \beta_i + (\eta_{j,t} + \varepsilon_{i,j,t}) + g(N_{i,j,t}^a) + \delta_{i,j,t} I\{C_{i,t-1} = j\} + \alpha_{i,j}$$

(14)

where $C_{i,t}$ is the choice variable of individual $i$ in period $t$; $I\{\cdot\}$ is the indicator function that gets the value of one if the statement in the parenthesis is true, and zero otherwise; and $\delta_{i,j,t}$ and $\alpha_{i,j}$ are parameters. We allow the state dependence effect to vary across consumers, products, and time. Its exact structure is presented in 5.2. Notice that the unobserved heterogeneity parameter $\alpha_{i,j}$ does not have an index $t$. Thus, it is common to all the shows offered by network $j$. Since it is unobserved by the researcher, ignoring it can bias the estimates of $\delta_{i,j,t}$.

Accordingly, the utility from the outside alternative is:

$$U_{i,J+1,t} = \gamma_{i,t} + (\eta_{J+1,t} + \varepsilon_{i,J+1,t}) + \delta_{i,\text{out},t} I\{C_{i,t-1} = J + 1\} + \alpha_{i,J+1}$$

(15)

5.1.2 Information Set and Expected Utility

As mentioned above, the television networks are multiproduct firms. We exploit this feature of the data in formulating the information set of consumers. The prior distribution of $\xi_{i,j,t}$ (which has an index $t$ since both $\eta_{j,t}$ and $X_{j,t}$ are time-dependent) is now:

$$\xi_{i,j,t} \sim N(\mu_{i,j}, \frac{1}{\sigma_{i,j}})$$

(16)

where, by definition, $\mu_{i,j} = E_t(\eta_{j,t}) + E_t(X_{j,t}) \beta_i$. In other words, we assume that although the individual is uncertain about product attributes, she knows the distribution of these attributes within each multiproduct firm. In a previous study using this data, we find empirical support for this assumption (Anand and Shachar, 2001a). Furthermore, the television industry serves Mankiw (1998) as a good example of the informational role of multiproduct firms. Referring to multiproduct firms as “brands”, he writes: “Establishing a brand name—and ensuring that it conveys the right information—is an important strategy for many businesses, including TV networks.”

It is worth noting that since we have only one week of data, and each show is offered only once during this week, the multiperiod aspect of our data is different from other consumer choice studies using panel data (for example, Eckstein, Horsky, and Raban 1988, and Crawford and Shum 31)

---

31 A New York Times article (September 20, 1996) that he cites, reads: “In television, an intrinsic part of branding is selecting shows that seem related and might appeal to a particular certain audience segment. It means ‘developing an overall packaging of the network to build a relationship with viewers, so they will come to expect certain things from us,’ said Alan Cohen, executive vice-president for the ABC-TV unit of the Walt Disney Company in New York.”
Their data follow consumer choices over multiple weeks and multiple purchase occasions. As a result, these studies focus on the dynamic learning of consumers through their experience with the product. Here, we account for these unobserved experiences through the miscellaneous signals, and focus on two other sources of information: advertising, and multiproduct firms.

The expected utility of the extended model is then:

\[ u_{i,j,t} = \left( \theta_{i,j,t} \mu_{i,j} + (1 - \theta_{i,j,t}) \xi_{i,j,t} \right) + \sigma_{i,j,t} z_{i,j,t} + g(N_{i,j,t}^a) + \varepsilon_{i,j,t} \]

\[ + \delta_{i,j,t} I\{C_{i,t-1} = j\} + \alpha_{i,j} \text{ for } j = 1, \ldots, J \]

where

\[ \theta_{i,j,t} \equiv \frac{\zeta_{i,j}^{\mu}}{\zeta_{i,j}^{a} + N_{i,j,t}^a \tau^a + N_{i,j,t}^m \tau^m} \quad \text{and} \quad \sigma_{i,j,t}^{\omega} \equiv \sqrt{\frac{\zeta_{i,j}^a N_{i,j,t}^a + \zeta_{i,j}^m N_{i,j,t}^m}{\zeta_{i,j}^{a} + N_{i,j,t}^a \tau^a + N_{i,j,t}^m \tau^m}} \]

The most significant difference between the expected utility of the extended model in (17) and the simple model in (10) is that \( \mu_i \) is replaced by \( \mu_{i,j} \), and \( \zeta_i^a \) is replaced by \( \zeta_{i,j}^a \). This means that product choices are a function of the distribution of product attributes within a multiproduct firm, not within the entire market. It follows that a consumer’s response to advertising signals is a function of \( \zeta_{i,j}^a \). The larger is \( \zeta_{i,j}^a \), the weaker is the consumer’s response. The logic is as follows: if the diversity of product attributes within a firm is small, the prior distribution is more precise, and advertising signals have a smaller effect on the posterior distribution and on choices. Since \( \zeta_{i,j}^a \) differs across firms, the informational effect of advertising should differ across them as well. This heterogeneity serves as an additional restriction in identifying the parameter of interest, \( \zeta^a \).

This result identifies two ways through which a firm provides consumers with information: (1) advertising, and (2) its product line. From the consumer’s point of view, advertising and the multiproduct firm’s characteristics are informational substitutes.

### 5.2 Functional forms

#### 5.2.1 Utilities

Having presented the data, we are ready to specify the utility more precisely. This specification introduces additional unobserved individual-specific parameters. For notational convenience, \( u_i \), whose density function is \( f_v \), includes all these parameters alongside those presented earlier in equation (11).
Attribute utility We restrict II and $f_v$ in equation (11) so that the match element in the utility is:

$$X_{j,t} \beta_i = \beta_{Gender} I\{\text{the gender of } i \text{ and the cast of } j, \text{tis the same} \}$$

$$+ \beta_{Age0} I\{\text{the age group of } i \text{ and the cast of } j, \text{tis the same} \}$$

$$+ \beta_{Age1} I\{\text{the distance between the age group of } i \text{ and the cast of } j, \text{tis one} \}$$

$$+ \beta_{Age2} I\{\text{the distance between the age group of } i \text{ and the cast of } j, \text{tis two} \}$$

$$+ \beta_{Family} I\{\text{i lives with her family and show } j, \text{tis about family matters} \}$$

$$+ \beta_{RaceIncome} I\{\text{one of the main characters in show } j, \text{tis African American} \}$$

$$+ Sitcom_{j,t}(\beta_{Sitcom} Y_i^\beta + v_i^{Sitcom})$$

$$+ ActionDrama_{j,t}(\beta_{AD} Y_i^\beta + v_i^{AD})$$

$$+ RomanticDrama_{j,t}(\beta_{RD} Y_i^\beta + v_i^{RD})$$

The first six $\beta$ parameters capture the effect of cast demographics on choices. As mentioned above, previous studies have demonstrated that viewers have a higher utility from shows whose cast demographics are similar to their own. Thus, one should expect to find that: (1) $\beta_{Age0} > \beta_{Age1} > \beta_{Age2}$, (2) $\beta_{Gender} > 0$, (3) $\beta_{Family} > 0$ and (4) $\beta_{Race} < 0$. We use an individual’s income as a proxy for her race, since information on race is not included in our data set.\textsuperscript{32} The taste for different show categories is a function of observed ($Y_i^\beta$) and unobserved ($v_i^{Sitcom}$, $v_i^{AD}$, $v_i^{RD}$) individual characteristics. The observed variables included in $Y_i^\beta$ are $Teens_i$, $GenerationX_i$, $BabyBoomer_i$, $Older_i$, $Female_i$, $Income_i$, $Education_i$, $Family_i$, and $Urban_i$. As mentioned earlier, these variables are defined in table 1. Each of these interactions between show category and individual characteristics is captured through a unique parameter. For example, the interaction between and Action Drama show and a female viewer is captured via $\beta_{Female}^{AD}$. All the other parameters are denoted accordingly.

Recall that the attribute utility is a function of the match element and $\eta_t$. While we can identify an $\eta_{j,t}$ for each combination of time slot and alternative (subject to one normalization), we prefer to fix this parameter for the duration of each show. Consequently, a half-hour show and a two hour movie each have one $\eta$ parameter. Given our intent in uncovering fundamental attributes

\textsuperscript{32}The proportion of African-Americans in the highest income category is disproportionately low, while it is disproportionately high in the lowest income category. This relationship persists for all income categories in between as well (U.S. Census Bureau 1995). Nielsen designed the sample to reflect the demographic composition of viewers nationwide and used 1990 Census data to achieve the desired result. We found that the income categories and the proportion of African-Americans in the Nielsen data closely match those in the U.S. population (National Reference Supplement 1995). Although our data set does not include information about race, Nielsen has it and reports its aggregate levels.
of the shows, this is a natural restriction.

State dependence parameters The utility presented in equation (14) includes the state dependence element, \( \delta_{i,j,t} I\{C_{i,t-1} = j\} \). Here we specify the structure of \( \delta_{i,j,t} \) and extend the state dependence to include another element. Specifically, we formulate the state dependence in the network utility as:

\[
\delta_{i,j,t} = \begin{cases} 
Y_i^\delta \delta^Y + X_{j,t}^\delta \delta^X + \nu_i^\delta \\
+ \delta_{First15} I\{The \ show \ on \ j started \ within \ the \ past \ 15 \ minutes\} \\
+ \delta_{Last15} I\{The \ show \ on \ j is \ at \ least \ one \ hour \ long \ and \ will \ end \ within \ 15 \ minutes\} \\
+ \delta_{Continuation} I\{The \ show \ on \ j started \ at \ least \ 15 \ minutes \ ago\}
\end{cases}
\]

where the observed variables included in \( Y_i^\delta \) are Teens\(_i\), Generation\(_X_i\), BabyBoomer\(_i\), Older\(_i\), Female\(_i\), Family\(_i\), and cable subscription status (Basic\(_i\) and Premium\(_i\)), and the vector \( X_{j,t}^\delta \) includes the following show categories: Sitcom\(_{j,t}\), ActionDrama\(_{j,t}\), RomanticDrama\(_{j,t}\), NewsMagazine\(_{j,t}\) and Sport\(_{j,t}\).\(^{33}\)

We allow the state dependence parameters to vary across age groups, gender, and family status because previous studies (Bellamy and Walker 1996) find preliminary evidence suggesting differences in the use of the remote control across these groups. We also allow these parameters to differ across individuals for unobserved reasons through \( \nu_i^\delta \).

These parameters can differ across show types as well. For example, one might expect \( \delta \) to be smaller for sports shows, since there is no clear plot in these shows when compared with dramas.

Finally, we allow the state dependence parameters to vary over time. Specifically, we expect \( \delta \) to be small in the first 15 minutes of a show, when viewers have not had enough time to get hooked by the show. For the same reason, the state dependence should be high during the last 15 minutes of a show. Furthermore, we expect that the state dependence is higher during a show than between shows (\( \delta_{Continuation} > 0 \)). Last, \( \delta_{InProgress} \) applies to individuals who were not watching network \( j \) in the previous time slot. Since the tendency to tune into a network to watch a show that has already been running for at least 15 minutes should be lower than for a show which has

\[^{33}\]The binary variable Basic\(_i\) is equal to one for the one third of the population that only has access to basic cable offerings, and the binary variable Premium\(_i\) is equal to one for the one third of the population that has both basic and premium cable offerings. The binary variable Sport\(_{j,t}\) is equal to one for the sport shows (Monday Night Football on ABC and Ice Wars on CBS), and the binary variable NewsMagazine\(_{j,t}\) is equal to one for news magazines (e.g., 48 Hours on CBS).
been on the air for less than 15 minutes, \( \delta_{\text{InProgress}} \) is expected to be negative.

We formulate the state dependence parameters in the outside utility as:

\[
\delta_{i,\text{out},t} I\{C_{i,t-1} = (J + 1)\} =
\begin{cases}
Y_i^\delta \delta^Y + v_i^\delta \\
+ \delta_{\text{Hour}} I\{\text{The time is either 9:00 PM or 10:00 PM}\} \\
+ \delta_{\text{FOX}10:00} I\{C_{i,t-1} = \text{Fox}\} I\{\text{The time is 10:00 PM}\}
\end{cases}
\]

Individual characteristics \( (Y_i^\delta \delta^Y + v_i^\delta) \) are included in exactly the same way for the outside alternative because they are meant to represent behavioral attributes intrinsic to individuals. Since the outside alternative includes the option to watch non-network shows, we allow the state dependence parameters to change “on the hour”. Notice that many non-network shows end on the hour, and thus we expect the \( \delta \) to be lower at that time (\( \delta_{\text{Hour}} < 0 \)). Furthermore, since Fox ends its national broadcasting at 10:00 PM our data cannot distinguish between viewers who stayed with Fox thereafter and those who chose the outside alternative. Thus, we expect \( \delta_{\text{FOX}10:00} \) to be positive.

**Outside alternative** We restrict \( \Pi \) and \( f_v \) in equation (11) to get \( \gamma_i = Y_i^\gamma \gamma \). Other than the standard observable individual characteristics (age, gender, income, education, family status, and area of residence), the vector \( Y_i^\gamma \) includes the variables \( \text{Basic}_i \), \( \text{Premium}_i \), \( \text{All}_i \), and \( \text{Same}_{i,t} \). The cable subscription status is included since the outside alternative includes the option of watching non-network shows—viewers with basic or premium cable have a larger variety of choices which can lead to a higher utility. The variable \( \text{All}_i \) (and \( \text{Same}_{i,t} \)) is equal to the average time that the individual watched television (and in the corresponding time slot \( t \)) during the previous days of the week. Individuals’ tendencies to watch television cannot be fully explained by their demographic characteristics. Thus, their prior viewing habits (\( \text{All}_i \), and \( \text{Same}_{i,t} \)) and the personal unobserved parameters \( \alpha_{i,J+1} \) are designed to capture other sources of such differences. Specifically, we estimate a 3x1 vector, \( \alpha_{i,J+1} \), where each element represents a different hour of the night. Since our data set starts on Monday, the variables \( \text{All}_i \), and \( \text{Same}_{i,t} \) have missing values for this day. Thus, we include specific parameters to account for this: \( \gamma_{\text{Monday}8:00} \) is added to the outside utility for the first hour on Monday night prime time, and \( \gamma_{\text{Monday}9:00} \) and \( \gamma_{\text{Monday}10:00} \) are defined analogously.

Finally, although we can, in principle, estimate \( \eta_{J+1,t} \) for each of the 60 time slots of the week, we impose the following restriction:

\[
\eta_{J+1,t} = \eta_{J+1,t+12} = \eta_{J+1,t+24} = \eta_{J+1,t+36} = \eta_{J+1,t+48} \quad \text{for } t = 1, \ldots, 12.
\]
This implies that the outside utility for the time slot between 8:00 and 8:15, for example, is the same across all the nights of the week. This still allows us to identify the expected increase in the outside utility during the night, but with 48 less parameters.

The persuasive effect  The functional form of $g(\cdot)$ is:

$$g(N_{i,j,t}^a) = [\rho_{1,MT,i} MondayTuesday_{j,t} + \rho_{1,WF,i} WednesdayFriday_{j,t}] N_{i,j,t}^a + [\rho_{2,MT,i} MondayTuesday_{j,t} + \rho_{2,WF,i} WednesdayFriday_{j,t}] (N_{i,j,t}^a)^2$$

where the binary variable $MondayTuesday_{j,t}$ is equal to one for shows which aired on Monday or Tuesday, and zero otherwise; and the binary variable $WednesdayFriday_{j,t}$ is equal to one for shows which aired on Wednesday, Thursday or Friday and is equal to zero otherwise. We allow the advertising parameters to differ across these two parts of the week to account for the problem of missing data mentioned above.

The quadratic term in exposures allows for a simple non-linear structure for the effect of $N_{i,j,t}^a$ on viewing decisions. When $\rho_{2,\cdot,i} < 0$, this non-linear structure represents the often termed “wear-out” effect of advertisements. Notice that we allow the persuasive parameters to differ across individuals for unobserved reasons.

5.2.2 Information Set

As mentioned above, the parameters of the prior distribution, $\mu_{i,j}$ and $\sigma_{i,j}^p$, depend on the distribution of product attributes of network $j$. In the estimation, we set the distribution of product attributes to be equal to the empirical distribution.

Furthermore, we restrict the prior distribution to account for a known strategy employed by the networks. Specifically, shows aired by the television networks between 10:00 and 11:00 PM tend to be dissimilar to those aired between 8:00 and 10:00 PM. For example, sitcoms are not broadcast after 10:00 PM on any night. Since this strategy is well-known, viewers are likely to have different prior beliefs about the scheduling for these two parts of the night. We account for that by allowing the prior distribution to differ not only across the networks, but also across the different parts of the night. Specifically, the prior distribution for each part of the night depends only on the distribution of the attributes of the shows which are broadcast during that part.\(^{34}\)

\(^{34}\)That is, for example, for shows aired between 8:00 to 10:00 PM, $\mu_{i,j}^{8-10} = \frac{1}{40} \sum_{t \in t^{8-10}} \xi_{i,j,t}$, where $t^{8-10}$ is the set of all the time slots between 8:00 and 10:00 PM during the week.
5.2.3 Density functions

We assume that the $\varepsilon_{i,j,t}$ are drawn from independent and identical Weibull (i.e., independent type I extreme value) distributions. As McFadden (1973) illustrates, under these conditions the viewing choice probability is multinomial logit.

The density function $f,$ is assumed to be discrete. Specifically, $v_i = v_k$ with probability $\frac{\exp(\lambda_k)}{\sum_{k=1}^K \exp(\lambda_k)}$ for all $k$. This means that we allow the population to be divided into $K$ different unobserved segments. The number of types $K$ is determined based on various information criteria.

5.2.4 Normalizations

Since there are five alternatives in each time slot, one can only identify four $\gamma$s. Thus, we normalize $\alpha_{k,ABC} = 0$ for each type $k$.

Since all the age categories are included in $Y_i^\beta$, one needs to normalize the $v_i^{Sitcom}, v_i^{AD}, v_i^{RD}$ for one of the types; therefore we set $v_i^{Sitcom} = v_i^{AD} = v_i^{RD} = 0$. Also, since all the age categories are included in $Y_i^\gamma$, one needs to normalize the $\gamma_{i,j+1}$ for one of the types, and $\eta_{j+1}$ for one of the time slots. We do so by setting $\alpha_{k=1,J+1,8:00}=\alpha_{k=1,J+1,9:00}=\alpha_{k=1,J+1,10:00} = 0$; and $\eta_{J+1,8:00} = 0$.

Similarly, since all the age categories are included in $Y_i^\delta$, we need to normalize the $v_i^\delta$ for one of the types. To do so, we set $v_i^\delta = 0$. Finally, since all the show categories are included in $X_{j,t}^\delta$, one of the parameters in $\delta^X$ needs to be normalized. We set $\delta_{NewsMagazine} = 0$.

Since all the age categories are included in $Y_i^\gamma$, one needs to normalize the $\eta_{j,t}$ of one of the shows. Notice that by increasing each of the $\eta_{j,t}$'s by 1, and each of the $\gamma$ parameters of the age categories by 1, the likelihood does not change. Thus, we set $\eta$ of Fox’s $X Files$ to be equal to zero.

Furthermore, since we estimate the $\eta_{j,t}$ for each show, we are required to normalize one of the $\beta_{Age}$ parameters, and the $\beta_{Sitcom}, \beta_{AD}$ and $\beta_{RD}$ for one age group. We do so by setting $\beta_{Age2} = \beta_{Sitcom} = \beta_{AD} = \beta_{RD} = 0$.

Finally, one cannot estimate all $K$ values of $\lambda_k$, which determine the size of the types. Therefore, we set $\lambda_{k=1} = 0$.

---

35The unobserved individual-specific parameters were introduced in sections 2 and 5. Together, the vector $v_i$ is $(\alpha_{i,ABC}, \alpha_{i,CBS}, \alpha_{i,NBC}, \alpha_{i,FOX}, \alpha_{i.OUT,8:00}, \alpha_{i.OUT,9:00}, \alpha_{i.OUT,10:00}, v_i^{Sitcom}, v_i^{AD}, v_i^{RD}, v_i^\delta, \rho_{1,MT,i}^1, \rho_{2,MT,i}^1, \rho_{1,WF,i}^1, \rho_{2,WF,i}^1, \alpha_{i,ABC}^m, \alpha_{i,CBS}^m, \alpha_{i,NBC}^m, \alpha_{i,FOX}^m)$. 

27
5.3 The likelihood function

The conditional choice probability is:

\[
f_1(C_{i,t} | C_{i,t-1}, W_{i,t}; v_i, z_{i,t}, \Omega) = \frac{\sum_{j=1}^{J+1} [I\{C_{i,t} = j\} \exp(\pi_{i,j,t}(C_{i,t-1}, W_{i,j,t}, v_i, z_{i,j,t}, \Omega))]}{\sum_{j=1}^{J+1} \exp(\pi_{i,j,t}(C_{i,t-1}, W_{i,j,t}, v_i, z_{i,j,t}, \Omega))}
\]  

(18)

where \(W_{i,j,t}\) is a vector of all the variables in the model (that is, \(X_{j,t}, Y_i, \) and \(N_{i,j,t}^{a}\)), \(W_{i,t}\) is the \(J\)-element vector whose \(j\)'th component is \(W_{i,j,t}\), \(z_{i,t}\) is the \(J\)-element vector whose \(j\)'th component is \(z_{i,j,t}\), \(\Omega\) is the vector of the parameters that are common to all the individuals,\(^{36}\) and \(\pi_{i,j,t} = u_{i,j,t} - \varepsilon_{i,j,t}\).

Let \(C_i = (C_{i,1}, \ldots, C_{i,T})\) denote individual \(i\)'s history of choices for the entire week. Although the \(\varepsilon_{i,j,t}\) are independent over time, the conditional probability of \(C_i\) is not simply the product of the conditional probability \(f_1(C_{i,t} | \cdot)\) for \(t = 1, \ldots, T\), for the following reason. For each individual, our panel includes twelve observations for each of the five nights of the week. Nielsen does not record viewing choices at 7:45 PM because between 7:00 PM and 8:00 PM, the affiliate stations broadcast local programming. Thus, the lagged choice for the first time slot of each night is missing.\(^{37}\) The history probability is then:

\[
f_2(C_i | W_i; v_i, z_i, \Omega) = \prod_{d=1}^{5} \left[ f_1(C_{i,(12d-11)} | W_{i,(12d-11)}; v_i, z_{i,(12d-11)}, \Omega) \prod_{t=(12d-10)}^{12d} f_1(C_{i,t} | C_{i,t-1}, W_{i,t}; v_i, z_{i,t}, \Omega) \right]
\]  

(19)

where \(W_i\) is the \(T\)-element vector whose \(t\)'th component is \(W_{i,t}\) and \(z_i\) is defined accordingly.

Notice that 8:00 PM is a natural starting point for the dynamic choice process of each night because the national networks do not air any programs between 7:00 PM and 8:00 PM. This means, for example, that the Boston affiliate station that airs ABC programming after 8:00 PM might broadcast at 7:45 PM a show that appears at the same time on the NBC affiliate in New York.

Integrating out the unobserved \(z\) of the first show on ABC, we get:

\[
\int_{\tilde{z}_1} f_2(C_i | W_i; v_i, (\tilde{z}_1, ..., \tilde{z}_{64}), \Omega) \phi(\tilde{z}_1) d\tilde{z}_1
\]

\(^{36}\)That is, \(\Omega = \{\eta_{j,t}, \beta_{\text{Gender}}, \beta_{\text{Age0}}, \beta_{\text{Age1}}, \beta_{\text{Age2}}, \beta_{\text{Family}}, \beta_{\text{Race}}, \beta_{\text{Sitcom}}, \beta_{\text{AD}}, \beta_{\text{RD}}, \delta^Y, \delta^X, \delta_{\text{First15}}, \delta_{\text{Last15}}, \delta_{\text{Continuation}}, \delta_{\text{InProgress}}, \delta_{\text{Hour}}, \delta_{\text{FOX10:00}}, \gamma, \epsilon^a\}.

\(^{37}\)Notice that the choice at 10:45 PM on the previous night is not the lagged choice for the 8:00 PM time slot.
where $\phi(\tilde{z}_1)$ is the standard normal density function. Repeating this integration for the other 63 shows in the week gives us the history probability unconditional on $z_i$, $f_3(C_i|W_i; v_i, \Omega)$. Recall that for any individual, $z_{i,j,t}$ is constant across all time slots of a specific show. In practice, because $z_{i,j,t}$ is show-specific, none of the integrals should include the entire history. Each integral includes only the time slots during which the relevant show is aired. For example, on Wednesday between 10:00 and 11:00 PM, each of the three major networks airs a one-hour show. Thus, for these time slots, the integration is only over three unobserved $z$’s. Indeed, the largest number of integrals for each time slot is 13. We use this feature to re-write the history probability in order to minimize the number of integrals for each time slot.

Finally, integrating out the unobserved individual-specific parameters, $v_i$, we get the marginal probability:

$$f_4(C_i|W_i; \Omega') = \sum_{k=1}^K f_3(C_i|W_i; v_k, \Omega) \frac{\exp(\lambda_k)}{\sum_{k=1}^K \exp(\lambda_k)}$$ (20)

where $\Omega'$ includes $\Omega$, the $v_k$’s, and the $\lambda$’s.

The likelihood function is:

$$L(\Omega') = \prod_{i=1}^I f_4(C_i|W_i; \Omega')$$ (21)

### 5.4 Simulating the marginal probability

Since $z_{i,j,t}$ is normally distributed, the integrals of $f_3(C_i|W_i; v_i, \Omega)$ do not have a closed form solution. Consistent and differentiable simulation estimators of $f_3(\cdot)$ and $f_4(\cdot)$ are

$$\hat{f}_3(C_i|W_i; v_i, \Omega) = \frac{1}{R} \sum_{r=1}^R f_2(C_i|W_i; v_i, z_r, \Omega)$$ (22)

and

$$\hat{f}_4(C_i|W_i; \Omega') = \sum_{k=1}^K \hat{f}_3(C_i|W_i; v_k, \Omega) \frac{\exp(\lambda_k)}{\sum_{k=1}^K \exp(\lambda_k)}$$ (23)

where the $z$’s are randomly drawn from the standard normal distribution. The Maximum Simulated Likelihood (MSL) estimator is then

$$\hat{\Omega}_{MSL}^{\prime} = \arg\max \sum_{i=1}^I \log \left( \hat{f}_4(C_i|W_i; \Omega') \right)$$ (24)

As explained in McFadden (1989) and Pakes and Pollard (1989), the $R$ variates for each individual’s $z$’s must be independent and remain constant throughout the estimation procedure. A drawback
of using MSL is the bias of $\hat{\Omega}'_{MSL}$ due to the logarithmic transformation of $f_3(\cdot)$. Despite this bias, the estimator obtained by MSL is consistent if $R \to \infty$ as $I \to \infty$, as detailed in Proposition 3 of Hajivassiliou and Ruud (1994). To attain negligible inconsistency, Hajivassiliou (1997) suggests increasing $R$ until the expectation of the score function is zero at $\hat{\Omega}'_{MSL}$. In our case this is achieved at $R = 400$.

In order to reduce the variance of $\hat{f}_3(\cdot)$, we employ importance sampling as described in the MC literature (see Rubinstein 1981). Our importance sampler is similar to the one used in BLP (1995). We draw the $z$‘s from a multivariate normal approximation of each person’s posterior distribution of $z$, given some preliminary MSL estimate of $\Omega'$, and appropriately weight the conditional probabilities to account for the oversampling from regions of $z$ which lead to higher probabilities of $i$‘s actual choices. For $R = 400$ we find that importance sampling reduces the RMSE of $\hat{f}_4(\cdot)$ to about 0.37 the size of the RMSE when not using importance sampling.

5.5 Identification

We start by considering the identification of a model under the assumption that the individual is fully informed (that is, under the assumption that $\frac{1}{\gamma_{i,j}} = 0$ for all $i$ and $j$). This discussion illustrates which parameters can be identified without the structural restrictions imposed by the full model and the additional variables introduced by this model.

5.5.1 Utility parameters

The parameters $\beta_{Gender}$, $\beta_{Age0}$, $\beta_{Age1}$, $\beta_{Age2}$, $\beta_{Family}$, $\beta_{Sitcom}$, $\beta_{AD}$, and $\beta_{RD}$ are identified by the correlation between $X_{i,j,t} \cdot Y_i^\beta$ and viewer choices. The unobserved tastes for show categories (the parameters $\nu_i^{Sitcom}$, $\nu_i^{AD}$, $\nu_i^{RD}$) are identified by the conditional viewer choice histories over show types. The parameter $\eta_{j,t}$ is identified by the aggregate show ratings, conditional on the show’s characteristics. The $\delta$ parameters are identified by the conditional state dependence—that is, by

38 We simulate all stochastic components of the model to construct an empirical distribution of the score function at $\hat{\Omega}'_{MSL}$. A quadratic form of this score function is asymptotically distributed $\chi^2$ with degrees of freedom equal to the number of parameters estimated.

39 The RMSE of $\hat{f}_4(C_tW_i;\Omega)$ is computed using $N_R$ sets of $R$ draws as:

$$RMSE(R) = \left[ \frac{1}{N_R} \sum_{n=1}^{N_R} \left( \frac{\hat{f}_4(C_tW_i;\Omega) - f_4,\text{true}}{f_4,\text{true}} \right)^2 \right]^{0.5}$$

where $f_4,\text{true}$ represents the true value. Since this true value is not computable, we evaluate $\hat{f}_4(\cdot)$ using $R = 2^{20}$ Monte Carlo draws and take this to be the true value.

Any reduction in the variance of the estimator for $\hat{f}_4(\cdot)$ reduces the bias and variance of the estimator of $\Omega'$. Quantifying the magnitude of this reduction is of interest. To our knowledge, constructing the empirical distribution of $\hat{\Omega}'_{MSL}$ via a bootstrapping method is the only way to proceed. Unfortunately, the cpu time required to compute $\hat{\Omega}'_{MSL}$ prohibits us from pursuing this goal.
the share of viewers who remain with an alternative over two sequential time slots, conditioning on $X_{j,t}$. Note that if there were no heterogeneity in show offerings by networks *within* a night, we cannot identify $\delta$. The parameter $\alpha_{i,j}$ is identified by the conditional viewer choice histories over networks. Notice that a positive $\alpha_{i,j}$ leads individual $i$ to view shows on network $j$ even when those shows do not fit her preferences well.\(^{40}\)

The conditional correlation between the number of advertising exposures and viewing choices identifies the persuasive parameters $\rho$. The targeting strategies of firms pose a challenge for the researcher in obtaining an unbiased estimate of $\rho$. Recall that table 2 provided an example of such targeting strategies. This implies that exposures to advertisements are not random. This problem is obviously not specific to our dataset. For example, commercials for beer appear frequently during sports broadcasts. The audience of these shows are likely consumers of beer anyway. Thus, differences in the aggregate beer consumption across people who have seen no commercial for beer versus those who have been exposed to several commercials need not reflect the persuasive power of advertising. Formally, this problem of endogeneity means that $E(\varepsilon_{i,j,t}|N^a_{i,j,t}) \neq E(\varepsilon_{i,j,t})$. Individual-level data and proper modeling of consumer preferences can resolve this problem.\(^{41}\) Cross-sectional individual-level data enables the researcher to estimate the observed heterogeneity of preferences, and with panel data one can even identify the unobserved individual-specific parameters. This is intended to account for consumer preferences so that what is left in $\varepsilon_{i,j,t}$ is not correlated with $N^a_{i,j,t}$. The success of this approach depends on firms’ abilities to assess consumer preferences. The only case where the endogeneity problem remains is if the firm can estimate preferences better than the researcher, and optimize its targeting strategy. Specifically, in our case it is worth noting that the television networks did not have access to individual-level panel data, like the one we obtained. Notwithstanding this, in section 6.1 we present further evidence suggesting that our modeling of consumer preferences is rich enough to resolve the endogeneity problem.

### 5.5.2 Information Set Parameters

The partial information model imposes some restrictions on the parameters and introduces new explanatory variables. These identify the information set parameters. We start by discussing the estimation of the prior distribution parameters, and then proceed to present the identification of the signals’ parameters.

\(^{40}\)As discussed in the literature, there are various sources of identifying $\delta$ separate from $\alpha$. See Chamberlain (1993) and Shachar (1994). The outside alternative provides us with an additional identifying source. When turning on the television, the individual’s “state” (lagged choices) does not attach her to any network. Thus, her viewing choice is influenced by $\alpha$ (and show characteristics), but not by $\delta$.

\(^{41}\)Such an approach is taken by Ackerberg (2001), Erdem and Keane (1996), and Shum (2000).
**Prior Distribution** Recall that the estimation of the prior distribution parameters was already discussed in sub-section 5.2.2. Once the parameters $\beta$ and $\eta$ are identified, we also have an estimate of all the variables which are a function of them, i.e., $\xi_{i,j,t}$, $\mu_{i,j}$ and $\xi^m_{i,j}$.

It is worth clarifying the identifying source of this distinction between $\alpha_{i,j}$ and $\mu_{i,j}$. In the model we set $\mu_{i,j} = 1T \sum_{t=1}^T \xi_{i,j,t} = 1T \sum_{t=1}^T \hat{\eta}_{j,t} + \left( 1T \sum_{t=1}^T X_{j,t} \right) \hat{\beta}_i$. Thus, the identification of $\mu_{i,j}$ is based on an explanatory variable that does not exist in a model where consumers are fully informed. Specifically, this variable is the mean offering of each network (for example, $1T \sum_{i,j} X_{j,t}$ for network $j$).

The following exercise might shed some additional light on the distinction between $\alpha_{i,j}$ and $\mu_{i,j}$. Assume that we estimate a model that does not include the distribution of product attributes of each multiproduct firm in the information set. Since we have a long panel for each viewer, we can estimate an individual–firm match parameter for each combination of individual and firm. We denote this “fixed effect” as $L_{i,j}$; note that it has 6700 different values (i.e., $J \cdot I$). After estimating $L_{i,j}$, one can run the following regression: $L_{i,j} = a_j + b \cdot \left( \frac{1}{T} \sum_{t=1}^T X_{j,t} \right) + \varepsilon_{L_{i,j}}^L$. The estimates of the parameters $a_j$ and $b$ can be used to calculate the predicted value of the “error” term, $\varepsilon_{L_{i,j}}^L$. One can then separate the observed individual–network match, $\hat{b} \cdot \left( \frac{1}{T} \sum_{t=1}^T X_{j,t} \right)$, from the unobserved match, $\varepsilon_{L_{i,j}}^L$. In our structural estimation, this separation between the observed match, $L_{i,j}$, and the unobserved match, $\alpha_{i,j}$, is built into the likelihood function, because the model includes the distribution of product attributes within each multiproduct firm in the information set.

**Product-specific signals** The parameters of the product-specific signals are $\zeta^m$ and $\zeta^a$. The sum of the precision of all the product-specific signals (advertising and miscellaneous), $\zeta^a N_{i,j,t}^a + \zeta^m_{i,j}$, enters $\pi_{i,j,t}$ (and thus the likelihood) only through $\theta_{i,j,t}$ and $\sigma^a_{i,j}$. Furthermore, neither $\zeta^a$ nor $\zeta^m$ enters the likelihood in any other form. Thus, we start by presenting the identification of $\zeta^a N_{i,j,t}^a + \zeta^m_{i,j}$, and then show how one can separately identify $\zeta^a$ and $\zeta^m_{i,j}$.

The dependence of $\theta_{i,j,t}$ and $\sigma^a_{i,j}$ on $\zeta^a N_{i,j,t}^a + \zeta^m_{i,j}$ lead to three identifying factors. The first is the effect of $\mu_{i,j}$ through $\theta_{i,j,t}$ on product choices. If $\zeta^a N_{i,j,t}^a + \zeta^m_{i,j} = 0$, then $\theta_{i,j,t} = 0$, and the choice of a product is not a function of $(\mu_{i,j} - \xi_{i,j,t})$. Thus, if the difference between $\mu_{i,j}$ and $\xi_{i,j,t}$ empirically affects the choices, then $\zeta^a N_{i,j,t}^a + \zeta^m_{i,j} > 0$. The larger the effect, the smaller is $\zeta^a N_{i,j,t}^a + \zeta^m_{i,j}$.

Notice that this source of identification relies on observed product and firm attributes.

---

42 Recall that one can rewrite equation (17) as:

$$u_{i,j,t} = \xi_{i,j,t} + \theta_{i,j,t}(\mu_{i,j} - \xi_{i,j,t}) + \sigma^a_{i,j} z_{i,j,t} + g(N_{i,j,t}^a) + \varepsilon_{i,j,t} + \delta_{i,j,t} I(C_{i,t-1} = j) + \alpha_{i,j} \quad \text{for } j = 1, \ldots, J$$

(26)
The other two identifying factors concern $\sigma_{i,j}^\omega$. Recall that $\sigma_{i,j}^\omega$ is a function of $\zeta N_{i,j,t}^a + \omega_{i,j}^m$. The variance $\sigma_{i,j}^\omega$ affects the observed consumer behavior in two ways. First, the larger is $\sigma_{i,j}^\omega$, the smaller is the correlation between any observed variables and choices. Second, the larger is $\sigma_{i,j}^\omega$, the stronger is the unexplained persistence in choices within a show.

Finally, $\zeta^a$ can be separately identified from $\omega_{i,j}^m$ using data on an individual’s exposure to advertisements, $N_{i,j,t}^a$. A positive $\zeta^a$ leads to a negative correlation between $N_{i,j,t}^a$ and our two measures of how ill-informed the individual is, $\theta_{i,j,t}$ and $\sigma_{i,j}^\omega$. Thus, if $\zeta^a > 0$, one would expect to find that an increase in $N_{i,j,t}^a$ would: (1) reduce the effect of $(\mu_{i,j} - \xi_{i,j,t})$ on choices; (2) increase the correlation between any observed variables and choices; and (3) decrease the unexplained persistence in choices. Recall that the first of these identifying factors reflects the implications presented in the model (that is, consumption-deterring and matchmaking). Specifically, since the expected utility is a positive function of $(\mu_{i,j} - \xi_{i,j,t})$, an increase in $N_{i,j,t}^a$ reduces the consumers’ tendency to purchase a product when $(\mu_{i,j} - \xi_{i,j,t}) > 0$ (and increases her tendency when $(\mu_{i,j} - \xi_{i,j,t}) < 0$).

Each of these identifying factors is different from that of the persuasive effect. As mentioned above, $\rho > 0$ implies that an increase in $N_{i,j,t}^a$ should have a positive effect on consumer $i$’s tendency to purchase alternative $j$ at time $t$. In contrast, $\zeta^a > 0$ implies that an increase in $N_{i,j,t}^a$ should have such a positive effect when $(\mu_{i,j} - \xi_{i,j,t}) < 0$, but a negative effect otherwise. This difference by itself enables us to identify both $\rho$ and $\zeta^a$. The dependence of $\sigma_{i,j}^\omega$ on $N_{i,j,t}^a$ further assists us in this task.

## 6 Results

Table 6 presents the estimates of the utility and the information set parameters (including the parameter of interest, $\zeta^a$).

The integrals in $f_3(C_i|W_i; \nu, \Omega)$ are evaluated numerically using importance sampling with 400 points from a pseudo-random sequence as detailed in Section 5.3. The (asymptotic) standard errors are derived from the inverse of the simulated information matrix.

The results are for a model with 5 segments ($K = 5$). The number of unobserved segments was determined by minimizing the Bayes Information Criterion (BIC). The largest segment consists of about 36% of the population, while the proportion of the smallest segment is about 11%. The sizes of the other segments are 0.22, 0.19 and 0.12; the $\lambda$ parameters, that determine the sizes of the segments, are reported in table 6a. The parameter of interest, $\zeta^a$, is reported last.

---

43 An easy way to think about this effect is by analogy with a simple regression model. Consider the case where $Y_i = \alpha + \beta X_i + \varepsilon_i$. We know that the larger the variance of $\varepsilon_i$, the smaller the observed correlation between $Y_i$ and $X_i$. The unobserved $\omega_{i,j,t}$ plays, in our model, a similar role to that of $\varepsilon_i$ in this simple regression.

44 The reported standard errors, therefore, neglect any additional variance due to simulation error in the numerical integration.
6.1 Utility parameters

A meaningful way to illustrate the effects of the parameters is in terms of probabilities. To do so, we define a baseline viewer as follows—a thirty year old female, who lives alone, in an urban area, has median income and education level, and is part of the largest unobserved segment.

6.1.1 Show attributes ($\beta$ and $\eta$)

The $\beta$ parameters are presented in table 6b and the $\eta$ parameters in table 6c.

As expected, viewers have heterogeneous preferences over show attributes (both cast demographics and show genres). They prefer shows whose cast demographic is similar to their own. The age of the cast has the largest effect. For example, the probability that the baseline viewer watches a show whose cast demographic matches hers (Generation X) is about three times larger than the probability if her age was 55 years. Additional findings are that (a) viewers living with their family like to watch shows about families, (b) viewers prefer shows whose cast is of the same gender as their own, and (c) the tendency to watch shows with an African-American cast decreases in income.

Viewers’ heterogeneous preferences over show genres depend on their observed and unobserved characteristics. Generation X viewers prefer romantic dramas over action dramas and sitcoms while viewers older than 35 years prefer action dramas over romantic dramas and sitcoms. Women prefer romantic dramas over the other show categories. People living in urban areas tend to like action dramas less than those living in rural areas. Finally, while high income viewers prefer sitcoms, low-income viewers prefer romantic dramas.

Shows are also different in their $\eta$ parameters. The show with the highest $\eta$ is the sitcom *Home Improvement* (ABC, Tuesday at 9:00 PM), and the one with the lowest is the generation-X romantic drama *Beverly Hills* (Fox, Monday at 9:00 PM).\(^{45}\)

6.1.2 State dependence ($\delta$)

The $\delta$ parameters are presented in table 6d.

The estimates are consistent with previous studies that found large state dependence in television viewing choices. For example, the average probability of watching the second time slot of a sitcom conditioned on watching the first is 81% in our sample. This probability depends both on the unobserved segments’ baseline state dependence ($\delta_{k=1} = 2.0873, \delta_{k=2} = 1.2737, \delta_{k=3} = 1.7934, \delta_{k=4} = 0.9320, \delta_{k=5} = 1.2772$) and the additional state dependence parameter specific to sitcoms ($\delta_{\text{sitcom}} = 0.7826$). The analogous probabilities for action dramas is 76%, for a romantic drama is

\(^{45}\)Notice that *Beverly Hills* was aired twice during this week: in its regular time slot (Wednesday at 8:00 PM) and in an irregular time slot (Monday at 9:00 PM). Its $\eta$ in the regular time slot is high.
81%, for sports shows is 53%, and for newsmagazines is 66%. State dependence is higher for shows with a plot line of some kind, as is the case with sitcoms and romantic dramas.

State dependence varies across viewers not only based on unobserved characteristics. It is lower for viewers with access to cable channels. State dependence in choices is, as commonly thought, higher for female viewers compared with males, although, somewhat surprisingly, there is not much difference across age groups.

Additional parameters help explain when viewers get hooked to a show. State dependence is highest in the last fifteen minutes of a drama, and lowest in the first fifteen minutes. Finally, as expected, switching into the middle of a show without watching its beginning is costly ($\delta_{\text{InProgress}} = -0.49$).

### 6.1.3 Individual-network match ($\alpha$)

The $\alpha$ parameters appear in table 6e. The largest segment of viewers is more likely to watch ABC and NBC than the other networks. The second largest segment likes NBC and Fox. The third largest segment prefers ABC over the other networks. The fourth largest segment does not have a strong preference for any of the networks, and the smallest segment likes CBS and dislikes Fox.

### 6.1.4 Persuasive Effect ($\rho$)

The $\rho$ parameters are presented in table 6f. It turns out that the two sets of parameters (for Monday and Tuesday, and for Wednesday through Friday) are quite similar. Therefore, we discuss only the Wednesday through Friday parameters.

While the effect on advertising differs across segments, the utility of each segment is a positive function of exposures to advertising. The effect is smallest for the largest viewer segment (segment 1). For this segment, the first exposure increases the probability of watching a show by 33% (from 6.2% to 8.2%), the second exposure increases the viewing probability by an additional 15%, and the effect of the third exposure is virtually zero. These results also illustrate the wear-out effect of advertising.

The differences in the $\rho$ parameters across the unobserved segments indicate that consumers’ responses to advertising are heterogeneous even if advertising is not part of the information set. Notice that these differences, unlike those resulting from the $\zeta^a$ parameter, do not depend on consumers’ tastes.\footnote{Table 7 provides evidence suggesting that our modeling of consumer preferences is rich enough to resolve the differences in the $\rho$ parameters across the unobserved segments.}
6.1.5 Preference for the outside alternative (γ)

The γ parameters are presented in table 6g.

There are clear differences in preferences for network television across individuals. Young viewers watch it the least and older viewers the most. Females, high-income viewers, and those who have access to cable watch less network television, but there is not much difference across education, family, and urban groups.

These demographic characteristics do not capture all the heterogeneity in viewer preferences for network television. The parameters $\alpha_{i,j+1}$, $\gamma_{all}$, and $\gamma_{same}$ capture the remaining heterogeneity.

Last, the utility from the outside alternative is found to increase during the evening.

6.2 Information Set Parameters

Individuals have three sources of information: (1) the distribution of product attributes within each multiproduct firm, (2) miscellaneous product-specific signals (word-of-mouth, media coverage, previous experience), and (3) advertising signals. The resulting parameters of the information set are: $\psi^{i,j}$ (the precision of the prior distribution), $\zeta^{m}_{i,j}$ (the precision of the miscellaneous signal), and $\zeta^{a}$ (the precision of advertising signals). The estimates are presented in this order below and in table 6h.

---

49 This may reflect the existence of popular cable channels for young viewers, like MTV and ESPN, and that young people have a higher utility from non television activities.

50 While this may seem surprising at first, recall that we only analyze prime time TV viewing habits here: it is likely that there might be significant differences across such viewer demographic groups in their tendency to watch daytime programs.

51 Since the variables All$_{i,j,t}$ and Same$_{i,j,t}$ have missing values on Monday (the beginning of the sample), the parameters $\gamma_{Mon, 8}$, $\gamma_{Mon, 9}$, and $\gamma_{Mon, 10}$ are all negative.

52 It seems from the estimates that the utility from the outside alternative between 8:15 and 10:45 is lower than the one at 8:00. This results from the bias of our $\gamma_{8,00}$ estimate. This parameter is biased, since we are missing the lagged choice of 7:45. Most viewers (about 75% of them) have their television off at 7:45, and thus they have a large state dependence effect. Since we do not include the lagged choice from 7:45, our estimate of $\gamma_{8,00}$ is upward biased. Recall, though, that this bias does not affect the parameters of interest in this study.
6.2.1 Precision of Prior distribution ($\zeta^\mu$)

Unlike $\zeta^m_{i,j}$ and $\zeta^a$, the prior distribution parameters $\zeta^\mu_{i,j}$ are not estimated directly but rather as a function of $\hat{n}_{j,t}$ and $\hat{\beta}_i$. The average $\zeta^\mu_{i,j}$ for each viewer segment and network are presented in the table. The average $\zeta^\mu_{i,j}$ is the highest for ABC and the lowest for Fox ($\zeta^\mu_{i,ABC} = 1.91$, $\zeta^\mu_{i,CBS} = 1.32$, $\zeta^\mu_{i,NBC} = 1.36$ and $\zeta^\mu_{i,Fox} = 0.54$). The finding for Fox may seem surprising since this network appears to offer the most homogeneous profile of shows: many GenerationX dramas, and no sitcoms or newsmagazines. However, recall that the attribute-utility is a function of both $\hat{n}_{j,t}$ and $X_{j,t}$. While the variance in $X_{Foz,t}$ is indeed the lowest among the four networks, the variance of $\hat{n}_{Foz,t}$ is the highest.

6.2.2 Precision of Miscellaneous Signal ($\zeta^m$)

The parameter $\zeta^m$ can be thought of as viewer $i$’s degree of familiarity with the shows on network $j$. The heterogeneity across individuals in their familiarity level is evident from the estimates ($\zeta^m_{i,j}$ varies from 0.034 to 14). Since these signals are product-specific, these estimates imply that individuals differ in their prior information about each show even without any exposures to advertising.

On average, viewers are more familiar with shows on ABC and NBC than those of the other networks (the averages are 4.95 for ABC, 3.58 for NBC, 0.81 for CBS and 0.21 for Fox). These estimates are sensible for the following reasons. The degree of familiarity with a network should be a positive function of (a) the ratings of its shows and (b) the “age” of its shows (i.e., the number of seasons that the shows were on the air). The reason is that information from both word-of-mouth sources and previous experience tends to be larger for successful and veteran shows. Even though NBC enjoyed the highest average rating (followed by ABC in second place) during the fall season of 1995, it was only third in the “ratings race” during the 1994 season (behind ABC and CBS). Moreover, while several of NBC’s highest rated shows in 1995 were in their first year of airing, the successful ABC shows were veterans. For example, one of ABC’s highest rated shows is Monday Night Football, which was in its 25th season. The low $\zeta^m_{i,j}$ for CBS and Fox are not surprising as well—their average rating lagged that of the other networks, and CBS had additionally introduced many new shows in the fall of 1995.

53 Specifically, since our estimate of the utility-attribute is $\hat{\xi}_{i,j,t} = \hat{n}_{j,t} + X_{j,t}\hat{\beta}_i$, it follows that (if we were not treating the two parts of the night, 8:00-10:00 PM and 10:00-11:00 PM, separately): $\zeta^\mu_{i,j} = \left[\frac{1}{T} \sum_t \left(\hat{\xi}_{i,j,t} - \frac{1}{T} \sum_t \hat{\xi}_{i,j,t}\right)^2\right]^{-1}$. We calculate $\zeta^\mu_{i,j}$ for each part of the night accordingly.
6.2.3 Precision of advertising signals (\(\zeta^a\))

We are now ready to discuss the parameter of interest, \(\zeta^a\). Contrary to previous studies that model advertising only as an element in the utility function, the model presented here also allows the information set to depend on advertising. Section 4 provided preliminary evidence to support this approach. In the structural estimation, the empirical evidence in favor of this approach rests on whether \(\zeta^a > 0\) or not. It turns out that the data support the theory—the estimate of \(\zeta^a\) is positive, statistically different from zero, and behaviorally important.

The estimate of \(\zeta^a\) is 1.172 with a standard error of 0.635. The Wald t-statistic is 1.846 while the one-sided critical t-value at the 5% significance level is 1.645. Table 8 illustrates that the likelihood is not symmetric around \(\zeta^a\). Based on a likelihood ratio test, even the model with \(\zeta^a\) set at 0.6 can be rejected against the model with \(\zeta^a\) free at the 1% significance level.\(^{54}\)

The effect of advertising through the information set is not just statistically significant, but behaviorally important as well. This is illustrated below in several ways.

Since the average \(\hat{\sigma}^m_{i,j}\) across viewers and networks is 2.4, and the average \(\hat{\sigma}^H_{i,j}\) is 1.28, the precision of two advertising signals is about the same as the precision of all other miscellaneous signals and the precision of a single advertising signal is equal to the precision of the prior distribution.

Two measures of how ill-informed an individual is were introduced in section 2—\(\theta_{i,j,t}\) and \(\sigma^\omega_{i,j,t}\). Tables 10a and 10b present these two measures as a function of the number of advertising exposures. On average, \(\theta = 0.5\) when the number of exposures is zero. With one exposure, \(\theta\) drops to 0.30. This table also illustrates the vast heterogeneity in viewers’ product-specific knowledge even without any advertising exposures (\(\theta\) ranges from 0.082 to 0.909). Furthermore, as expected, the effect of an advertising signal is lower for viewers who are more informed. For example, for the fifth segment, \(\theta_{Fox}\) drops from 0.909 to 0.321 with exposure to a single advertisement, while for the third segment, \(\theta_{ABC}\) hardly changes with advertising exposures. Since advertising signals inform individuals about product attributes, each additional signal increases the informativeness level of the individual. Thus, the effect of the \(n\)'th signal is smaller than the effect of \((n-1)\)'th. For example, while the average \(\theta\) drops from 0.5 to 0.30 with the first exposure, it drops to 0.22 with the second exposure.

\(^{54}\)There is no assurance that our estimates yield a global maximum of the likelihood given the large number of parameters and unobserved segments. In order to decrease the prospect of a local maximum, the model was estimated several times, with a different variable being dropped each time. In all these cases, the estimate of \(\zeta^a\) was positive and statistically different from zero.

The identification discussion of \(\rho\) (section 5.5.1) and the evidence in footnote 47 suggest that \(\hat{\rho}\) is unbiased. It is worth noting that even if \(\hat{\rho}\) were biased, this should not affect our estimate of \(\zeta^a\). First, as noted in section 5.5.2, the identifying sources of \(\zeta^a\) are different from those of \(\rho\). Specifically, \(\rho\) is identified by the correlation between advertising exposures and choices. \(\zeta^a\) is identified by (a) the correlation between the heterogeneous responses to advertising exposures and consumers’ heterogeneous preferences over product attributes, and (b) the dependence of \(\sigma^\omega_{i,j}\) on \(N^a_{i,j,t}\). Second, table 9 provides evidence that the correlation between \(\hat{\rho}\) and \(\hat{\zeta^a}\) is close to 0.
Section 7 presents additional ways to assess the informative effect of advertising.

6.3 Goodness of fit

The fit of the model in each of the sixty time slots is tested using the $\chi^2$ test in Heckman (1984) which applies to models with parameters estimated from micro data. Constructing a single $\chi^2$ statistic to test the model is not computationally feasible. The test statistic is a quadratic form of the difference between observed cell counts and expected cell counts using the model. We fail to reject the null hypothesis that the model is correctly specified for only 9 of the 60 time slots using a 5% significance level.

7 Applications

This section illustrates the normative and positive consequences of informative advertising based on the structural estimates. It starts with several normative implications such as the consumption-deterring aspect of advertising and its matchmaking role. The positive consequences center around targeting strategies of firms.

7.1 Matching

Section 4 provided some preliminary evidence that (a) a consumer’s response to advertising depends on her taste for the product and (b) exposure to advertising improves the matching of consumers and products. Reexamining these implications using the structural estimates reinforces these findings, as described below.

Consumption deterring role The model implies that exposure to advertising increases the tendency to purchase when $(\mu_{i,j} - \xi_{i,j,t})$ is negative and decreases this propensity when $(\mu_{i,j} - \xi_{i,j,t})$ is positive. The columns in table 11 differ with respect to the sign of $(\mu_{i,j} - \xi_{i,j,t})$ while the rows differ with respect to the number of exposures (the table ignores observations in which the television was turned off). The structural estimates enable us to identify the sign of $(\mu_{i,j} - \xi_{i,j,t})$ for each combination of individual and show. The task is complicated by the variation of tastes across the unobserved segments. Thus, for each combination of individual and show, we calculate the probability that $(\mu_{i,j} - \xi_{i,j,t})$ is positive. The numbers in each cell of table 11 record the viewing choices of individuals weighted by the probabilities defined above. Consider the following example. An individual was not exposed to any advertisement for each of the shows in a specific time slot.

55 The number of cells that fully partition the response vector space is $5^{60}$ (when ignoring the absence of FOX in the 10:00-11:00 PM time slots). This test is also a special case in Andrews (1998).

39
She watched the show on ABC, for which the probability that \( \mu_{i,j} - \xi_{i,j,t} \) is positive is 0.25. If this were the only observation, the first row of the table (i.e., zero exposures) would have had a 0.25 in the left column and 0.75 in the right, and all the other rows would have had zeros. We repeat this calculation for all time slots for this individual and then for all individuals, and average all these numbers to get table 11. As before, we equalize the persuasive effect across all alternatives by focusing only on those observations where individuals were exposed to the same number of advertisements for all the competing shows in a time slot.

The first row indicates that even without any advertising exposures, individuals are somewhat informed about show characteristics (otherwise both probabilities in this row should have been close to 0.5). As expected, an increase in the number of advertisements increases the difference between the two columns. In other words, the purchase tendency increases when \( \mu_{i,j} - \xi_{i,j,t} \) is negative and falls when \( \mu_{i,j} - \xi_{i,j,t} \) is positive. Specifically, when comparing choices with no advertising exposures for each show (first row) versus one advertising exposure (second row), we find that the tendency to watch shows which are not a good fit decreases from 0.438 to 0.314. In contrast, the propensity to choose shows which are a good fit increases from 0.561 to 0.685. The difference between the first two rows is statistically significant at the 1% significance level, whereas the difference between the last two rows is not.

Detecting the consumption-deterring aspect of advertising requires precise estimates of consumer preferences. This requirement makes a non-structural examination difficult. Using the structural estimates, table 11 reveals the strength of the consumption-deterring aspect in the data. Notice that the choices and the exposures to advertising in this table are based on the actual data, not on predictions.

A similar finding appears in the second part of the table, 11 (b), which includes only observations in which the television was on both in the current time slot and the previous one. In the third and fourth parts of the table, 11(c) and 11(d), we also ignore cases where \( \mu_{i,j} - \xi_{i,j,t} \) is close to zero. As expected, the difference between the two columns is stronger.

**Matching of consumers with products** Since advertising guides consumers to products that better fit their tastes, it improves the match of consumers and products. Next, we assess the magnitude of the matching effect in terms of the changes in both consumers’ choices and their utility. The assessment is based on simulating consumers’ choices under two scenarios for the number of exposures. In the first, the number of advertising exposures is equal to its value in the data set, \( N_{i,j,t}^a \), and in the second it is equal to \( N_{i,j,t}^a + 1 \). The columns of table 12 present these two scenarios.

The first panel of table 12 describes the fit between individuals and the alternatives they chose
for all the time slots.\textsuperscript{56} While individuals do not know for certain which is their best alternative, the researcher is not uncertain about product attributes and can calculate the alternative that yields the highest utility for each individual. It turns out that although individuals are uncertain about product attributes, they choose their best alternative in about 95\% of the cases for all the exposure scenarios in the first panel. The reason for this high percentage is the large magnitude of the state dependence parameter. The increase in the fit between consumers and products as a result of an exposure to an additional advertisement is small when focusing on the percent of first-best choices (from 94.67\% to 95.56\% and from 94.35\% to 95.21\%) but large when viewed from the perspective of non-first best choices (a decrease from 5.33\% to 4.44\% and from 5.65\% to 4.79\%).\textsuperscript{57} This means that the number of cases in which individuals do not choose their first-best alternative decreases by about 16\%.

The value of informative advertising can be better assessed by focusing on times when individuals depart from the status quo, in other words when they switch. The second panel serves this purpose, by presenting the fit (a) only for observations where the television was turned off in period \( t - 1 \) but was on in period \( t \), and (b) where \( \delta_{i,j,t} = 0 \). Indeed, removing the effect of state dependence expresses itself in the smaller percent of first-best choices compared to the first panel. Furthermore, the effect of advertising is also larger than in the first panel. Specifically, while the percent of first-best choices is 32.03 in our dataset, it is 55.37 for the scenario \( N^a_{i,j,t} + 1 \). This also means that the percent of non-first best choices falls from 67.97 to 44.63, a decrease of 34\%.

Both panels include the average utility experienced by individuals, \( \bar{U} \equiv \frac{1}{TT} \sum_{i=1}^{I} \sum_{t=1}^{T} \sum_{j=1}^{J+1} \sum_{j=1}^{J+1} [U_{i,j,t} I\{C_{i,t}^{\text{sim}} = j\}] \), where \( C_{i,t}^{\text{sim}} \) are the simulated choices. As discussed in the model section (2.4) and in the preliminary evidence (section 4), one would expect this realized utility to increase with the number of advertising exposures \( N^a_{i,j,t} \). The result in table 12 supports this view. The change in utility is 0.0176 and 0.0217 in the first panel, and 0.159 in the second panel. Since \( \gamma_{\text{Basic}} = 0.2 \) per time slot, an increase in utility of 0.159 as a result of an additional exposure to one advertisement for each show equals 79.5\% of the increase in utility from having a cable connection.

### 7.2 Targeting Strategies

In recent years there has been a surge in the number of media outlets such as newspapers, television channels, magazines, websites, etc. For example, between 1985 and 1995, the number of significant cable channels increased from 31 to 87.\textsuperscript{58} Similarly, between 1988 and 2001, the number of magazines wholly devoted to particular areas of interest such as “nursing” and “fish and fisheries” grew

\textsuperscript{56}The results are obtained using 100 simulation draws (of the product-specific signals and \( \varepsilon_{i,j,t} \)) for each individual and setting the persuasive effect to be equal to \( g(N^a_{i,j,t}) \) for both scenarios.

\textsuperscript{57}All these changes are statistically significant at the 1\% level.

\textsuperscript{58}Parsons and Frieden (1998).
from 16 and 19, respectively, to 135 and 64. Advertisers thus face increasingly segmented audiences. Segmentation provides firms an opportunity to improve the effectiveness of advertisements as long as they can identify the segment whose response to their advertisements is the strongest. The model presented here might assist in this task. It tracks the effects of advertising through both the utility and the information set, and as a result provides a precise estimate of each consumer’s response to advertising.

This subsection evaluates the power of this approach in targeting segmented audiences. This is done first by comparing the targeting strategies employed by the television networks with strategies suggested by our model as being “optimal”. Then, the actual placements of advertisements is compared with the optimal ones. In order to get a sensible comparison between the actual and the optimal placements, we compute the optimal strategies subject to two constraints. These constraints follow particular decisions made by the networks. First, the number of advertisements for each show in the exercise is set to be equal to the actual number. Second, the maximum number of advertisements for show $j$ in show $k$ (other than in a show that lasted two hours) is set to be one.

The mechanism by which the optimal placement is chosen is as follows: (1) an advertisement for a show is hypothetically removed from the schedule, and replaced by an advertisement for the same show in the first time slot of the week. The total predicted rating of the network is then calculated. (2) Step 1 is repeated, this time by placing the advertisement in the second time slot, and then in every subsequent one until the time slot that precedes the show in question. (3) The advertisement in question is placed in the time slot that yields the highest predicted rating. (4) Steps (1)-(3) are repeated for every advertisement in the schedule, holding the previous advertisements in their “optimal” locations. (5) This entire cycle (steps 1 through 4) is repeated until no ad location changes through an entire cycle. Another mechanism is to calculate the optimal placements is to change the locations of advertisements for all the shows simultaneously rather than sequentially. However, since the computation of this alternative mechanism is infeasibly complex, we employ the mechanism described above. Obviously, this means that the chosen mechanism only provides a lower bound on rating improvements.60

The logit models in table 13 present the first comparison between actual and “optimal” strategies. There are two dependent variables, $d_{j,k}^a$ and $d_{j,k}^p$. The binary variable $d_{j,k}^a = 1$ if an advertisement for show $j$ appeared in show $k$, and is equal to zero otherwise. The binary variable $d_{j,k}^p = 1$ if it is optimal (based on the model) to place an advertisement for show $j$ in show $k$, and zero otherwise.

The first four independent variables in the table are measures of match between show $j$ and $k$.

59 http://www.magazine.org/resources/fact_sheets/ed7_8_01.html

60 While the network executive maximizes profits, we ignore cost considerations and focus on ratings instead of revenues. Goettler (1998) shows that this approximation is not costly when computing an optimal show schedule.
(Demographic_Match\(_{j,k}\), Genre_Match\(_{j,k}\), Same_Hour\(_{j,k}\), and Preceding_Show\(_{j,k}\)), and the last variable (Rating\(_k\)) measures the ratings of show \(k\). All these variables are formally defined in the table. The variables Demographic_Match\(_{j,k}\) and Genre_Match\(_{j,k}\) are based on the similarity in the attributes of shows \(j\) and \(k\). The binary variables Same_Hour\(_{j,k}\) and Preceding_Show\(_{j,k}\) are equal to one if shows \(j\) and \(k\) are broadcast on the same hour of a night, and if show \(k\) directly precedes show \(j\), respectively. Last, Rating\(_k\) is equal to the number of people that watched show \(k\) in our dataset. The independent variables are not based on theory, but rather on the model estimates and intuition. One should expect the variables Demographic_Match\(_{j,k}\) and Genre_Match\(_{j,k}\) to have a positive effect on the probabilities of the events \(d^a_{j,k} = 1\) and \(d^p_{j,k} = 1\) because of the informative role of advertising. Specifically, the model implies that a consumer’s response to advertising are higher when the product better fits her tastes. Accordingly, the coefficient on Same_Hour\(_{j,k}\) is expected to be positive because individuals tend to watch television in regular time slots. A similar argument based on state dependence suggests that the coefficient on Preceding_Show\(_{j,k}\) should be positive. The expected sign of the coefficient on Rating\(_k\) is positive when the dependent variable is \(d^p_{j,k}\) since show \(k\) provides a larger number of individuals who would be exposed to the advertisement. When the dependent variable is \(d^a_{j,k}\), however, the expected sign is ambiguous since larger ratings also imply higher opportunity costs.

In general, the coefficients have the expected signs and the actual strategies are consistent with the optimal ones. Specifically, the effect of the match variables (Demographic_Match\(_{j,k}\) and Genre_Match\(_{j,k}\)) on \(d^p_{j,k}\) are positive and statistically different from zero at the 1% level. In other words, it is optimal to place advertisements for show \(j\) in show \(k\) when the two shows have similar attributes. Furthermore, the coefficients are quite similar in the logit model with \(d^a_{j,k}\) as a dependent variable, implying that the networks usually follow this optimal strategy.

The coefficient of Preceding_Show\(_{j,k}\) is positive in both models. This means that it is optimal to place advertisements in the preceding show and that the networks apply this strategy. As expected, there is a difference in the effect of ratings (Rating\(_k\)) between the two models. Interestingly, the coefficient on Same_Hour\(_{j,k}\) is not statistically different from zero even at the 10% level in the best response model.

Another way to evaluate how close the actual strategies are to the optimal ones is by comparing the placements themselves. The percent of advertisements that are placed in the optimal locations is 59.9%. Furthermore, the model predicts that if the networks were to replace their locations with the optimal ones, their ratings would have increased by only 3.9%. (Specifically, the market share of ABC would have increased from 0.0762 to 0.0807, of CBS from 0.0565 to 0.0586, of NBC from 0.0874 to 0.0908, and of Fox from 0.0487 to 0.0497.) Given that our exercise does not account for the cost of placing advertisements, these results indicate that the network strategies
are quite close to the optimal ones.\footnote{Recall as well that optimal strategies are constrained to follow certain decisions made by the networks. When these constraints are released, the rating gains are higher than those reported in the text.}

We have also solved for a Nash equilibrium as follows. For any network $A$, its best response is calculated as described above. Given this new schedule of advertising for $A$ and the actual schedule for $C$ and $D$, the best response of network $B$ is now computed. The process is repeated until a full cycle of the four networks does not yield any change in the schedule. The percent of advertisements that are placed in the optimal locations is 59.1%. The model predicts that if the networks were to replace their locations with the equilibrium placements, their ratings would have increased by only 2.8%. (Specifically, the market share of ABC would have increased from 0.0777 to 0.0812, of CBS from 0.0565 to 0.0579, of NBC from 0.0874 to 0.09, and of Fox from 0.0487 to 0.0492.)

8 Conclusion

The findings in this study are relevant for both theoretical and empirical work. Here, we highlight potential extensions and applications of these findings.

A vast literature in economics studies how firms strategically reveal information to consumers through signaling. But consumers do not always have to infer information from firms’ behavior. This study shows that consumers rely on advertising content to obtain information about product attributes. Advertising, of course, is not the only simple way in which firms inform consumers. Automobiles produced by Volvo, for example, are perceived by consumers to be safe, and movies produced by Disney are perceived by viewers to be family-friendly. These multiproduct firms and others provide information by selecting a clear product line. Elsewhere we show that the information set of consumers indeed depends on the profile of multiproduct firms (Anand and Shachar 2001a). These findings, together, raise several questions that have not been addressed thus far. For example, what does this theory imply about the relationship between advertising intensity and product diversity in equilibrium? Is the predicted relationship observed in the data? And what are the consequences of the informational role of both advertising and product line choices for consumer welfare?

The finding that advertising can deter consumption raises new issues concerning the targeting strategies of firms. The danger of exposing the wrong consumer to advertisements encourages firms to invest in precise targeting and increases the demand for media channels that deliver highly segmented audiences. Consumers might also learn about product attributes from firms’ targeting strategies. The equilibrium of markets with rational consumers, informative advertising, and segmented media channels is the focus of Anand and Shachar (2001b). It is shown there that, in equilibrium, rational consumers respond positively to advertising intensity. This implies that the
effect of advertising intensity on the utility (captured in the model here by $\rho$) may be a consequence of rational behavior.

The increased demand for segmented media channels, coupled with technological advances, has created new services (e.g., TiVo) that enable firms to target consumers individually. The theoretical and empirical consequences of informative advertising and individualized targeting services merit attention.

The growing importance of advertising has led researchers to include advertising intensity as an independent variable in their demand estimation. These models do not allow the information set to depend on advertising. The findings in this paper suggest that these models might be misspecified. Suggesting that each one of these models would employ exactly the same approach taken here is unreasonable due to the complexity of such a task. This raises the need to develop a simple modeling approach that can proxy the effect of advertising through the information set. For example, one approach is to allow $\rho$ to be a function of consumer tastes.

While the evidence presented in section 4 illustrate simple ways to assess the informative role of advertising, additional uncomplicated examinations can be fruitful. For example, many firms allow consumers to return products after purchase. Data on exposures to advertising combined with information on product returns can be used to reveal the informative role of advertising. Finally, while we have outlined the advantages of the dataset created for this study, such data is likely to have a caveat. Specifically, advertisements for television shows can be thought of as product samples and as a result may be more informative than advertising for other products. Examining the robustness of our results would require creating similar datasets for other products.
References

[1] Ackerberg, Daniel. (2001). “Advertising, Learning, and Consumer Choice in Experience Good Markets: A Structural Empirical Examination.” Mimeo.

[2] Anand, Bharat and Ron Shachar (2001a) “Multiproduct Firms, Information, and Loyalty”, Harvard Business School Working Paper No. 00-069.

[3] Anand, Bharat and Ron Shachar (2001b) “An Anti-Groucho Marx Theory of Advertising: Targeting as a Signal of Product Match,” Mimeo, http://www.tau.ac.il/~rroonn/Papers/Marx.pdf.

[4] Anderson, Simon and Stephen Coate (2000), “Market Provision of Public Goods: The Case of Broadcasting,” Working Paper.

[5] Andrews, Donald. (1998). “Chi-Square Diagnostic Tests for Econometric Models: Introduction and Applications,” Journal of Econometrics, 37, 135-156.

[6] Becker Gary and Kevin Murphy (1993) “A Simple Theory of Advertising as a Good or Bad,” Quarterly Journal of Economics, 108, 941-964.

[7] Bellamy, Robert and James Walker (1996), Television and the Remote Control. New York: The Guilford Press.

[8] Berry, Steven, James Levinsohn, and Ariel Pakes (1995), “Automobile Prices in Market Equilibrium, Econometrica, 63, 841-890.

[9] Bond, Jonathan and Richard Kirshenbaum (1998), Under the Radar: Talking to Today’s Cynical Consumer. New York: John Wiley & Sons.

[10] Butters, G. (1977) “Equilibrium Distribution of Prices and Advertising,” Review of Economic Studies, 44, 465-492.

[11] Chamberlain, Gary (1993) “Multivariate Regression Models for Panel Data” in Maddala, G. S., ed. The econometrics of panel data. Volume 1. Aldershot, U.K. pages 266-307.

[12] Crawford, Gregory and Matthew Shum (2000) “Uncertainty and Learning in Pharmaceutical Demand,” Mimeo.

[13] DeGroot, M. (1989) Probability and Statistics, (2nd edition) Addison Wesley.

[14] Eckstein, Zvi, Dan Horksy, and Yoel. Raban (1988). “An Empirical Dynamic Model of Optimal Brand Choice”. Foerder Institute of Economic Research, Working Paper no. 88.

[15] Erdem, Tulin and Michael Keane (1996) “Decision-Making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets,” Marketing Science, 15 (1), 1-20.

[16] Goettler, Ronald (1999) “Advertising Rates, Audience Composition, and Competition in the Network Television Industry,” GSA Working Paper no. 1999-E28, Department of Economics, Carnegie Mellon University.

[17] Goettler, Ronald and Ron Shachar (2001) “Spatial competition in the network television industry,” RAND Journal of Economics, 32 (4), 624-656.

[18] Grossman, Gene and Carl Shapiro (1984) “Informative Advertising with Differentiated Products,” Review of Economic Studies, 51, 63-82.

[19] Hajivassiliou, Vassilis (1997), “Some Practical Issues in Maximum Simulated Likelihood”, London School of Economics Discussion Papers Series EM/97/340.
[20] Hajivassiliou, Vassilis and Paul Ruud (1994), “Classical Estimation Methods for LDV Models Using Simulation,” in Handbook of Econometrics, Volume IV, ed. by R.F. Engel and D.L. McFadden, Amsterdam: Elsevier Science.

[21] Heckman, James (1984) “The $\chi^2$ Goodness of Fit Statistic for Models with Parameters Estimated from Microdata.” Econometrica, 52, 1543-1548.

[22] Jacoby, Jacob and Wayne Hoyer (1982) “Viewer Miscomprehension of Televised Communication: Selected Findings,” Journal of Marketing, 46 (4), 12-26.

[23] Jacoby, Jacob and Wayne Hoyer (1989) ”The Comprehension/Miscomprehension of Print Communication: Selected Findings,” Journal of Consumer Research, 15(4), 434-443.

[24] Mahajan, Vijay, Eitan Muller, and Jerry Wind (2000) New Product Diffusion Models, Kluwer Academic Publishers, Boston, Massachusetts.

[25] Lancaster, Kelvin (1971) Consumer Demand: A New Approach. New York: Columbia University Press.

[26] Mankiw, Gregory (1998) Principles of Economics, Harcourt College Publishers.

[27] McFadden, Daniel (1989), “A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration,” Econometrica, 57, 995–1026.

[28] McFadden, Daniel (1981) “Econometric Models of Probabilistic Choice, ” in Structural Analysis of Discrete Data with Econometric Applications, ed. by Charles Manski and Daniel McFadden. Cambridge, MIT Press.

[29] McFadden, Daniel (1973), “Conditional Logit Analysis of Qualitative Choice Behavior,” in Frontiers in Econometrics, ed. by P. Zarembka, New York: Academic Press.

[30] Milgrom Paul and John Roberts (1986) “Price and Advertising Signals of Product Quality,” Journal of Political Economy, 94, 796-821.

[31] National Reference Supplement (1995-96). Nielsen Media Research.

[32] Nevo, Aviv (2001) “Measuring Market Power in the Ready-to-Eat Cereal Industry”, Econometrica 69(2), 307-342.

[33] Pakes, Ariel and Pollard, David (1989), “Simulation and the Asymptotics of Optimization Estimators,” Econometrica, 57, 1027–1057.

[34] Parsons and Frieden (1998). The Cable and Satellite Television Industries. Boston: Allyn and Bacon, 349-351.

[35] Rubinstein, R. (1981), Simulation and the Monte Carlo Method, New York: John Wiley & Sons, Inc.

[36] Rust, Roland and Mark Alpert (1984), “An Audience Flow Model of Television Viewing Choice,” Marketing Science, 3, 113–124.

[37] Salop, Steven (1979) “Monopolistic Competition with Outside Goods,” The Bell Journal of Economics, 10 (1), 141-156.

[38] Shachar, Ron and John Emerson (2000) “Cast Demographics, Unobserved Segments, and Heterogeneous Switching Costs in a TV Viewing Choice Model,” Journal of Marketing Research, 37 (2), 173-186.

[39] Shachar, Ron and Bharat Anand (1998) “The Effectiveness and Targeting of Television Advertising.” Journal of Economics and Management Strategy, 7 (3), 363-396.

[40] Shachar, Ron (1994) “A Diagnostic Test for the Sources of Persistence in Individuals’ Decisions,” Economics Letters, 45, 7-13.
[41] Shum, Matthew (1999) “Advertising and Switching Behavior in the Breakfast Cereals market,” Mimeo.

[42] Tirole, Jean (1988) *The Theory of Industrial Organization*, The MIT Press, Cambridge, Massachusetts.

[43] U.S. Bureau of the Census (1995).
## Figure 1

### Television Schedule, November 6-10, 1995

| Day   | Network | 8:00               | 8:30               | 9:00               | 9:30               | 10:00               | 10:30               |
|-------|---------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| Mon.  | ABC     | The Marshal        | Pro Football: Philadelphia at Dallas |
|       | CBS     | The Nanny          | Can’t Hurry Love   | Murphy Brown       | High Society       | Chicago Hope       |
|       | NBC     | Fresh Prince of Bel-Air | In the House     | Movie: She Fought Alone |
|       | FOX     | Melrose Place      | Beverly Hills 90210 | Affiliate Programming: News |
| Tue.  | ABC     | Roseanne           | Hudson Street      | Home Improvement   | Coach              | NYPD Blue          |
|       | CBS     | The Client         | Movie: Nothing Lasts Forever |
|       | NBC     | Wings              | News Radio         | Frasier            | Pursuit of Happiness | Dateline NBC |
|       | FOX     | Movie: Bram Stoker’s Dracula | Affiliate Programming: News |
| Wed.  | ABC     | Ellen              | The Drew Carey Show | Grace Under Fire   | The Naked Truth    | Prime Time Live    |
|       | CBS     | Bless This House   | Dave’s World       | Central Park West  | Courthouse         |
|       | NBC     | Seaquest 2032      | Dateline NBC       | Law & Order        |
|       | FOX     | Beverly Hills 90210 | Party of Five      | Affiliate Programming: News |
| Thu.  | ABC     | Movie: Columbo: It’s All in the Game | Murder One |
|       | CBS     | Murder, She Wrote  | New York News      | 48 Hours           |
|       | NBC     | Friends            | The Single Guy     | Seinfeld           | Caroline in the City | E.R.             |
|       | FOX     | Living Single      | The Crew           | New York Undercover | Affiliate Programming: News |
| Fri.  | ABC     | Family Matters     | Boy Meets World    | Step by Step       | Hangin’ With Mr. Cooper | 20/20 |
|       | CBS     | Here Comes the Bride | Ice Wars: USA vs. The World |
|       | NBC     | Unsolved Mysteries | Dateline NBC       | Homicide: Life on the Street |
|       | FOX     | Strange Luck       | X-Files            | Affiliate Programming: News |
Table 1
Individual Observable Characteristics: Definitions and Summary Statistics

| Variable | Definition | Mean     | Standard Deviation |
|----------|------------|----------|--------------------|
| Teens    | Viewer is between 6 and 17 years old (in November 1995) | 0.1421 | 0.3491 |
| Gen – X  | Viewer is between 18 and 34 years old (in November 1995) | 0.2400 | 0.4272 |
| Boom     | Viewer is between 35 and 49 years old (in November 1995) | 0.2764 | 0.4474 |
| Older    | Viewer is older than 50 years | 0.3415 | 0.4742 |
| Female   | Female viewer | 0.5319 | 0.4991 |
| Male     | Male viewer | 0.4681 | 0.4991 |
| Family   | Viewer lives in a household with (according to Nielsen codes) a “woman of the house” (i.e., female over the age of 18) present | 0.4304 | 0.4953 |
| Income   | Measured on unit interval, where the limits are: zero if the income is less than $10,000, and one if the income is $40,000 and over | 0.8333 | 0.2259 |
| Education| Measured on unit interval, where the limits are: zero if the years of school are less than 8, and one if the it is 4 or more years college | 0.7421 | 0.2216 |
| Urban    | Viewer lives in one of the 25 largest cities in U.S | 0.4149 | 0.4929 |
| Basic    | Viewer has basic cable service | 0.3642 | 0.4813 |
| Premium  | Viewer has basic and premium cable service | 0.3588 | 0.4798 |

Table 2
Advertising deters consumption: Effect of exposure to advertising on the propensity to view 48 Hours

| Number of exposures to advertisements on each of the networks | "Low" Match | "High" Match |
|-------------------------------------------------------------|-------------|--------------|
| 0 or 1                                                       | 5.9% (2164) | 16.7% (96)   |
| 2 or more                                                   | 4.3% (36)   | 28.0% (68)   |

Notes:
1. The numbers in each cell denote the percentage of viewers who watched 48 hours. The numbers in parentheses are the total number of individuals within that cell.
2. "Low" Match indicates that the viewer watched newsmagazines less than 10% of the time that these shows were aired during the rest of the week (i.e., Dateline NBC on Tuesday 10 PM, Wednesday 9 PM, and Friday 9 PM, Prime Time Live on Wednesday 10 PM, and 20/20 on Friday 10 PM). "High" Match indicates that the viewer watched these newsmagazines at least 10% of the time.
| Variable          | Definition                                                                                                                                                                                                 | Coefficient | Standard Error |
|-------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------|----------------|
| $N_{i,k}^a$       | The number of advertisements for show $k$ that individual $i$ was exposed to. (It is equal also to the number of advertisements for each show in the same time slot that individual $i$ was exposed to).                           | 16.8059     | 4.9204         |
| $\text{NewsMatch}_{i,k}$ | The number of time slots that individual $i$ watched a news-magazine (other than show $k$) divided by the number of time slots during which a news-magazine (other than show $k$) was aired. | 5.6266      | 1.2303         |
| $\text{Information}_{i,k}$ | $N_{i,k}^a \cdot (1 - \text{NewsMatch}_{i,k})$                                                                                                           | -16.5208    | 5.7109         |
| $\text{Leadin}_{i,k}$ | A binary variable that is equal to one if individual $i$ was watching the network that airs show $k$ in the proceeding time slot, and equal to zero otherwise.                                             | 2.3429      | 1.1568         |
| $\text{Same}_{i,k}$ | The average number of time slots that individual $i$ watched TV at the same time that show $k$ is aired during all the days of the week (excluding the day that show $k$ is aired). | 1.5967      | 0.7116         |
| $\text{All}_{i,k}$ | The average number of time slots that individual $i$ watched TV during all the days of the week (excluding the day that show $k$ is aired).                                                                  | 0.1621      | 0.8069         |
| $20/20_k$        | A fixed effect for the show 20/20                                                                                                                                                                         | -0.6391     | 0.1855         |
| $\text{Dateline}_k$ | A fixed effect for the show Dateline NBC                                                                                                                                                                  | -0.7282     | 0.1793         |
| $\text{Constant}$ | ---                                                                                                                                                                                                   | -13.1080    | 0.1396         |

**Note:** The number of observations is 2075.
### Table 4a
**Advertising and Matching:**
Effect of exposure to advertising on the chosen “match” between viewers and shows

| Number of exposures to advertisements for each network | Average | Standard deviation | Number of observations |
|--------------------------------------------------------|---------|--------------------|------------------------|
| 0                                                      | 0.3679  | 0.1757             | 3581                   |
| 1 or more                                              | 0.3914  | 0.1517             | 177                    |

**Note:**
1. The chosen match is equal to \( \sum_{j=1}^{J} I\{C_i = j\} \cdot Match_{i,j} \) where \( Match_{i,j} \) is the demographic match between viewers and shows. This variable is based on three demographic characteristics: age, gender, and family status. It counts the number of characteristics that are identical for both the show and the individual. For example, for a Generation-X single female viewer and a Generation-X show with a single, male cast, \( Match_{ij} = 2 \). For additional details see footnote 27 in the text.
2. The observations in this table are for individuals who did not watch any of the networks in the previous time slot.

### Table 4b
**Advertising and Matching:**
Effect of exposure to advertising on the chosen “match” between viewers and shows

| Number of exposures to advertisements for each of the networks ABC and CBS | Individuals who did not watch ABC or CBS in previous time slot | Individuals who watched ABC in previous time slot | Individuals who watched CBS in previous time slot |
|--------------------------------------------------------------------------|---------------------------------------------------------------|--------------------------------------------------|--------------------------------------------------|
| 0                                                                        | Average 0.5317                                               | 0.4751                                           | 0.5691                                           |
|                                                                          | Std. Deviation 0.2028                                        | 0.1976                                           | 0.1977                                           |
|                                                                          | Number of obs 1184                                           | 1271                                             | 773                                              |
| 1                                                                        | Average 0.5120                                               | 0.5529                                           | 0.5477                                           |
|                                                                          | Std. Deviation 0.1820                                        | 0.1930                                           | 0.1815                                           |
|                                                                          | Number of obs 72                                             | 225                                              | 108                                              |
| 2 or more                                                                | Average 0.5544                                               | 0.5959                                           | 0.5929                                           |
|                                                                          | Std. Deviation 0.1687                                        | 0.1585                                           | 0.1826                                           |
|                                                                          | Number of obs 13                                             | 60                                               | 23                                               |
### Table 4c
**Advertising and Matching:**
Effect of exposure to advertising on the chosen “match” between viewers and shows

| Number of exposures to advertisements for each of the networks ABC and NBC | Individuals who did not watch ABC or NBC in previous time slot | Individuals who watched ABC in previous time slot | Individuals who watched NBC in previous time slot |
|---|---|---|---|
| 0 | Average | 0.5385 | 0.5732 | 0.5247 |
|  | Std. Deviation | 0.2031 | 0.2021 | 0.1903 |
|  | Number of obs | 1291 | 1316 | 840 |
| 1 | Average | 0.5177 | 0.5031 | 0.5666 |
|  | Std. Deviation | 0.2037 | 0.1901 | 0.1896 |
|  | Number of obs | 90 | 169 | 256 |
| 2 or more | Average | 0.5856 | 0.5021 | 0.5843 |
|  | Std. Deviation | 0.1467 | 0.1431 | 0.1717 |
|  | Number of obs | 15 | 32 | 85 |

### Table 4d
**Advertising and Matching:**
Effect of exposure to advertising on the chosen “match” between viewers and shows

| Number of exposures to advertisements for each of the networks CBS and NBC | Individuals who did not watch CBS or NBC in previous time slot | Individuals who watched CBS in previous time slot | Individuals who watched NBC in previous time slot |
|---|---|---|---|
| 0 | Average | 0.5341 | 0.5824 | 0.4967 |
|  | Std. Deviation | 0.2106 | 0.2104 | 0.1953 |
|  | Number of obs | 1103 | 678 | 850 |
| 1 | Average | 0.5268 | 0.5213 | 0.5462 |
|  | Std. Deviation | 0.1857 | 0.1997 | 0.2020 |
|  | Number of obs | 84 | 123 | 214 |
| 2 or more | Average | 0.5735 | 0.5478 | 0.5938 |
|  | Std. Deviation | 0.2032 | 0.1699 | 0.1723 |
|  | Number of obs | 28 | 61 | 113 |
### Table 5
Advertising and Repeat Purchase: Probit estimates of effect of advertising exposures on the propensity to switch from a show

| Variable   | Definition                                                                 | Coefficient | Standard Error |
|------------|---------------------------------------------------------------------------|-------------|----------------|
| $N_{i,k}$  | The number of advertisements for show $k$ that individual $i$ was exposed to | -0.0217     | 0.0054         |
| $All_{i,k}$ | The average number of time slots that individual $i$ watched TV during all the days of the week (excluding the day that show $k$ is aired). | -0.0095     | 0.0071         |
| Constant   | ---                                                                       | 0.2562      | 0.0169         |

**Notes:**
1. The dependent variable is binary, and is equal to one if the individual switched away from a show that she was watching earlier, and zero otherwise.
2. The number of observations is 2665

### Table 6a
Structural Estimates: the sizes of the segments

| Parameter | Parameter | Standard Error | Size of the segment |
|-----------|-----------|----------------|---------------------|
| $\lambda_{k=1}$ | 0         | ---            | 0.3646              |
| $\lambda_{k=2}$ | -1.1925   | 0.1988         | 0.1106              |
| $\lambda_{k=3}$ | -1.1379   | 0.2146         | 0.1168              |
| $\lambda_{k=4}$ | -0.5258   | 0.1649         | 0.2155              |
| $\lambda_{k=5}$ | -0.6388   | 0.2335         | 0.1925              |
| Parameter          | Estimate | Standard Error | Parameter          | Estimate | Standard Error |
|--------------------|----------|---------------|--------------------|----------|---------------|
| $\beta_{\text{Gender}}$ | 0.2207   | 0.0438        | $\beta_{\text{Teens}}$ | 0.0000   | ----          |
| $\beta_{\text{Age0}}$ | 1.1038   | 0.0956        | $\beta_{\text{GenX}}$ | -0.2543  | 0.2482        |
| $\beta_{\text{Age1}}$ | 0.7977   | 0.0753        | $\beta_{\text{BabyBoomer}}$ | -0.3948  | 0.2363        |
| $\beta_{\text{Age2}}$ | 0.0000   | ----          | $\beta_{\text{Older}}$ | -0.4102  | 0.2506        |
| $\beta_{\text{Family}}$ | 0.3628   | 0.1125        | $\beta_{\text{Female}}$ | 0.7535   | 0.1256        |
| $\beta_{\text{Race}}$ | -0.8338  | 0.2513        | $\beta_{\text{Income}}$ | -1.4195  | 0.2696        |
| $\beta_{\text{Teens Sitcom}}$ | 0.0000   | ----          | $\beta_{\text{Education}}$ | -0.6431  | 0.2800        |
| $\beta_{\text{GenX Sitcom}}$ | -0.7701  | 0.1800        | $\beta_{\text{Family}}$ | 0.1810   | 0.1328        |
| $\beta_{\text{BabyBoomer Sitcom}}$ | -0.8261  | 0.1781        | $\beta_{\text{Urban}}$ | 0.0556   | 0.1146        |
| $\beta_{\text{Older Sitcom}}$ | -1.1824  | 0.1926        | $\beta_{\text{k=1 Sitcom}}$ | 0.0000   | ----          |
| $\beta_{\text{Female Sitcom}}$ | 0.3756   | 0.0818        | $\beta_{\text{k=1 AD}}$ | 0.0000   | ----          |
| $\beta_{\text{Income Sitcom}}$ | -0.2850  | 0.2146        | $\beta_{\text{k=1 RD}}$ | 0.0000   | ----          |
| $\beta_{\text{Education Sitcom}}$ | -0.2578  | 0.2068        | $\beta_{\text{k=2 Sitcom}}$ | 0.8258   | 0.2451        |
| $\beta_{\text{Family Sitcom}}$ | 0.1857   | 0.1075        | $\beta_{\text{k=2 AD}}$ | -0.4012  | 0.2110        |
| $\beta_{\text{Urban Sitcom}}$ | -0.0490  | 0.0819        | $\beta_{\text{k=2 RD}}$ | -2.2048  | 0.4688        |
| $\beta_{\text{Teens AD}}$ | 0.0000   | ----          | $\beta_{\text{k=3 Sitcom}}$ | 0.4971   | 0.2532        |
| $\beta_{\text{GenX AD}}$ | -0.4530  | 0.1873        | $\beta_{\text{k=3 AD}}$ | 1.3646   | 0.2300        |
| $\beta_{\text{BabyBoomer AD}}$ | -0.3628  | 0.1832        | $\beta_{\text{k=3 RD}}$ | 0.4368   | 0.3361        |
| $\beta_{\text{Older AD}}$ | -0.1960  | 0.1902        | $\beta_{\text{k=3 Sitcom}}$ | -0.5281  | 0.2106        |
| $\beta_{\text{Female AD}}$ | 0.4009   | 0.0848        | $\beta_{\text{k=4 Sitcom}}$ | -0.2880  | 0.1959        |
| $\beta_{\text{Income AD}}$ | -0.6648  | 0.2321        | $\beta_{\text{k=4 AD}}$ | -0.0913  | 0.2649        |
| $\beta_{\text{Education AD}}$ | -0.0526  | 0.2074        | $\beta_{\text{k=5 Sitcom}}$ | -1.0587  | 0.3402        |
| $\beta_{\text{Family AD}}$ | 0.0094   | 0.1011        | $\beta_{\text{k=5 AD}}$ | 0.0329   | 0.2973        |
| $\beta_{\text{Urban AD}}$ | -0.1525  | 0.0879        | $\beta_{\text{k=5 RD}}$ | -0.1440  | 0.4272        |
| Show name                  | Estimate | Standard Error |
|----------------------------|----------|----------------|
| *The Marshal*              | -2.8790  | 0.4857         |
| *Monday Night Football*    | -0.7858  | 0.4660         |
| *Roseanne*                | 0.2503   | 0.4037         |
| *Hudson Street*           | -0.3880  | 0.4233         |
| *Home Improvement*        | 1.5279   | 0.3979         |
| *Coach*                   | -0.2780  | 0.3897         |
| *NYPD Blue*               | -0.0980  | 0.3493         |
| *Ellen*                   | -0.3270  | 0.4061         |
| *The Drew Carey Show*     | -0.4180  | 0.4187         |
| *Grace Under Fire*        | 0.2358   | 0.3656         |
| *The Naked Truth*         | -0.4921  | 0.3947         |
| *Prime Time Live*         | -0.3662  | 0.4420         |
| *Columbo*                 | -0.5516  | 0.3223         |
| *Murder One*              | -1.7448  | 0.3765         |
| *Family Matters*          | 1.2251   | 0.4465         |
| *Boy Meets World*         | 0.2548   | 0.4357         |
| *Step by Step*            | 0.0264   | 0.3797         |
| *Hangin' With Mr. Cooper*| 0.8526   | 0.4200         |
| *20/20*                   | -0.0328  | 0.4308         |

| Show name                  | Estimate | Standard Error |
|----------------------------|----------|----------------|
| *The Nanny*                | -0.9715  | 0.6057         |
| *Can't Hurry Love*         | -1.4918  | 0.6663         |
| *Murphy Brown*             | -2.4282  | 0.6220         |
| *High Society*             | -2.6447  | 0.6959         |
| *Chicago Hope*             | -0.2969  | 0.6263         |
| *The Client*               | -0.8840  | 0.3882         |
| *Nothing Lasts Forever*    | 0.6917   | 0.3907         |
| *Bless This House*         | 0.0083   | 0.4187         |
| *Dave's World*             | 0.7447   | 0.4443         |
| *Central Park West*        | -0.8127  | 0.3986         |
| *Courthouse*               | -1.3956  | 0.4507         |
| *Murder, She Wrote*        | -0.1609  | 0.3365         |
| *New York News*            | -1.4036  | 0.3757         |

| Show name                  | Estimate | Standard Error |
|----------------------------|----------|----------------|
| *CBS*                      |          |                |
| *48 Hours*                 | -0.7520  | 0.4892         |
| *Here Comes The Bride*     | -0.3655  | 0.4363         |
| *Ice Wars*                 | -0.3727  | 0.3895         |
| *Fresh Prince*             | -0.1552  | 0.5400         |
| *In The House*             | -0.4798  | 0.5681         |
| *She Fought Alone*         | -0.9509  | 0.4271         |
| *Wings*                    | 0.0105   | 0.4008         |
| *NewsRadio*                | -0.1312  | 0.4305         |
| *Frasier*                  | 0.2670   | 0.3785         |
| *Pursuit of Happiness*     | -1.3123  | 0.4257         |
| *Dateline NBC (T)*         | -0.2772  | 0.4537         |
| *Seaquest 2032*            | -1.3784  | 0.3738         |
| *Dateline NBC (W)*         | -0.3429  | 0.3962         |
| *Law and Order*            | -0.8912  | 0.3449         |
| *Friends*                  | 0.7674   | 0.3888         |
| *The Single Guy*           | 0.1014   | 0.3967         |
| *Seinfeld*                 | 0.5041   | 0.3680         |
| *Caroline in the City*     | -0.2557  | 0.3803         |
| *E.R.*                     | 0.7088   | 0.3145         |
| *Unsolved Mysteries*       | -0.2550  | 0.3550         |
| *Dateline NBC (F)*         | -0.7417  | 0.4040         |
| *Homicide*                 | -1.4604  | 0.3640         |
| *Melrose Place*            | -1.2749  | 0.8488         |
| *Beverly Hills 90210 (M)*  | -3.3298  | 0.7978         |
| *Bram Stoker's Dracula*    | -1.2685  | 0.4060         |
| *Beverly Hills 90210 (W)*  | 1.3548   | 0.3856         |
| *Party of Five*            | -1.1786  | 0.4072         |
| *Living Single*            | 0.2984   | 0.4739         |
| *The Crew*                 | 0.6113   | 0.5035         |
| *New York Undercover*      | -0.3484  | 0.4713         |
| *Strange Luck*             | -1.0986  | 0.4136         |
| *X-Files*                  | 0.0000   | ----           |

| Show name                  | Estimate | Standard Error |
|----------------------------|----------|----------------|
| *FOX*                      |          |                |
| *Murder, She Wrote*        | -0.1609  | 0.3365         |
| *New York News*            | -1.4036  | 0.3757         |

Table 6c
Structural Estimates: the $\eta$ Parameters
| Parameter           | Estimate | Standard Error |
|---------------------|----------|----------------|
| $\delta_{Siccom}$   | 0.7826   | 0.0959         |
| $\delta_{ActionDrama}$ | 0.5266   | 0.0995         |
| $\delta_{RomanticDrama}$ | 0.7392   | 0.1031         |
| $\delta_{NewsMagazine}$ | 0.0000   | ----           |
| $\delta_{Sport}$    | -0.5558  | 0.1144         |
| $\delta_{k=1}$      | 2.0876   | 0.1275         |
| $\delta_{k=2}$      | 1.2739   | 0.1266         |
| $\delta_{k=3}$      | 1.7936   | 0.1391         |
| $\delta_{k=4}$      | 0.9326   | 0.1191         |
| $\delta_{k=5}$      | 1.2773   | 0.1423         |
| $\delta_{Basic}$    | -0.3147  | 0.0438         |
| $\delta_{Premium}$  | -0.2139  | 0.0463         |
| $\delta_{Female}$   | 0.1157   | 0.0359         |
| $\delta_{Family}$   | -0.0223  | 0.0456         |
| $\delta_{Teens}$    | 0.0000   | ----           |
| $\delta_{GenerationX}$ | -0.0344  | 0.0712         |
| $\delta_{BabyBoomer}$ | -0.0486  | 0.0643         |
| $\delta_{Older}$    | -0.0171  | 0.0737         |
| $\delta_{Continuation}$ | 1.9578  | 0.0949         |
| $\delta_{Out}$      | 1.1590   | 0.1051         |
| $\delta_{First15}$  | -0.5493  | 0.0782         |
| $\delta_{Last15}$   | 0.4852   | 0.1281         |
| $\delta_{Hour}$     | -0.2732  | 0.0868         |
| $\delta_{FOX10:00}$ | 0.7644   | 0.1539         |
| $\delta_{InProgress}$ | -0.4920  | 0.0722         |
| Viewer Segment | Parameter | Estimate | Standard Error |
|----------------|-----------|----------|----------------|
| 1              | \( \alpha_{ABC} \) | 0        | ----           |
|                | \( \alpha_{CBS} \) | 0        | ----           |
|                | \( \alpha_{NBC} \) | 0        | ----           |
|                | \( \alpha_{FOX} \) | 0        | ----           |
| 2              | \( \alpha_{ABC} \) | 0.3666   | 0.1545         |
|                | \( \alpha_{CBS} \) | -0.0148  | 0.1357         |
|                | \( \alpha_{NBC} \) | -0.5551  | 0.3225         |
| 3              | \( \alpha_{ABC} \) | 0.4676   | 0.1801         |
|                | \( \alpha_{CBS} \) | 0.3865   | 0.1455         |
|                | \( \alpha_{NBC} \) | 0.2172   | 0.2488         |
| 4              | \( \alpha_{ABC} \) | 0        | ----           |
|                | \( \alpha_{CBS} \) | -0.0222  | 0.1323         |
|                | \( \alpha_{NBC} \) | 0.3008   | 0.1173         |
|                | \( \alpha_{FOX} \) | 0.1474   | 0.1817         |
| 5              | \( \alpha_{ABC} \) | 0        | ----           |
|                | \( \alpha_{CBS} \) | -0.0960  | 0.2238         |
|                | \( \alpha_{NBC} \) | -0.0744  | 0.2029         |
|                | \( \alpha_{FOX} \) | -0.3102  | 0.2984         |
| Viewer Segment | Parameter | Estimate  | Standard Error |
|---------------|-----------|-----------|----------------|
| 1             | $\rho_{1,MT}$ | 0.3097    | 0.1308         |
|               | $\rho_{2,MT}$ | -0.0836   | 0.0403         |
|               | $\rho_{1,WF}$ | 0.3865    | 0.0909         |
|               | $\rho_{2,WF}$ | -0.0777   | 0.0271         |
| 2             | $\rho_{1,MT}$ | 0.1159    | 0.2461         |
|               | $\rho_{2,MT}$ | 0.0163    | 0.0749         |
|               | $\rho_{1,WF}$ | 0.1722    | 0.1541         |
|               | $\rho_{2,WF}$ | 0.0044    | 0.0453         |
| 3             | $\rho_{1,MT}$ | 0.4767    | 0.2366         |
|               | $\rho_{2,MT}$ | -0.0542   | 0.0697         |
|               | $\rho_{1,WF}$ | 0.346     | 0.1598         |
|               | $\rho_{2,WF}$ | -0.0504   | 0.0464         |
| 4             | $\rho_{1,MT}$ | 0.4717    | 0.2013         |
|               | $\rho_{2,MT}$ | -0.1124   | 0.0716         |
|               | $\rho_{1,WF}$ | 0.2183    | 0.1223         |
|               | $\rho_{2,WF}$ | -0.0140   | 0.0385         |
| 5             | $\rho_{1,MT}$ | 1.2772    | 0.2695         |
|               | $\rho_{2,MT}$ | -0.1882   | 0.0794         |
|               | $\rho_{1,WF}$ | 1.0071    | 0.1460         |
|               | $\rho_{2,WF}$ | -0.0801   | 0.0326         |
| Parameter       | Estimate | Standard Error | Parameter       | Estimate | Standard Error |
|-----------------|----------|----------------|-----------------|----------|----------------|
| $\gamma_{Basic}$ | 0.2033   | 0.0387         | $\gamma_{Monday:9:00}$ | -1.5725  | 0.2065         |
| $\gamma_{Premium}$ | 0.2860   | 0.0409         | $\gamma_{Monday:10:00}$ | -1.1502  | 0.1648         |
| $\gamma_{All}$  | -0.5756  | 0.0985         | $\gamma_{8:00}$  | 0.0000   | ----           |
| $\gamma_{Same}$ | -0.6717  | 0.0615         | $\gamma_{8:15}$  | -1.1088  | 0.0972         |
| $\gamma_{Teens}$| 3.1675   | 0.3410         | $\gamma_{8:30}$  | -0.7968  | 0.0993         |
| $\gamma_{GenerationX}$ | 2.6041 | 0.3663         | $\gamma_{8:45}$  | -0.9490  | 0.1302         |
| $\gamma_{BabyBoomer}$ | 2.4314 | 0.3693         | $\gamma_{9:00}$  | -0.9607  | 0.1366         |
| $\gamma_{Older}$ | 1.9424   | 0.3720         | $\gamma_{9:15}$  | -0.9166  | 0.1571         |
| $\gamma_{Female}$ | 0.1338   | 0.0637         | $\gamma_{9:30}$  | -0.8597  | 0.1517         |
| $\gamma_{Income}$ | -0.4199  | 0.1645         | $\gamma_{9:45}$  | -0.7489  | 0.1666         |
| $\gamma_{Education}$ | -0.0925 | 0.1504         | $\gamma_{10:00}$ | -0.5399  | 0.2649         |
| $\gamma_{Family}$ | 0.1248   | 0.0794         | $\gamma_{10:15}$ | -0.4485  | 0.2780         |
| $\gamma_{Urban}$ | -0.0289  | 0.0619         | $\gamma_{10:30}$ | -0.0856  | 0.2770         |
| $\gamma_{Monday:8:00}$ | -1.3008 | 0.2024         | $\gamma_{10:45}$ | -0.5399  | 0.2649         |

| $\alpha_{k=1,Out:8-9PM}$ | 0.0000   | ---- | $\alpha_{k=3,Out:10-11PM}$ | -1.8062  | 0.2240         |
| $\alpha_{k=1,Out:9-10PM}$ | 0.0000   | ---- | $\alpha_{k=4,Out:8-9PM}$ | -0.0908  | 0.2051         |
| $\alpha_{k=1,Out:10-11PM}$ | 0.0000   | ---- | $\alpha_{k=4,Out:9-10PM}$ | -0.1366  | 0.1361         |
| $\alpha_{k=2,Out:8-9PM}$ | -0.2169  | 0.2562 | $\alpha_{k=4,Out:10-11PM}$ | -0.3435  | 0.1843         |
| $\alpha_{k=2,Out:9-10PM}$ | 0.3043   | 0.1643 | $\alpha_{k=5,Out:8-9PM}$ | 0.0123   | 0.3154         |
| $\alpha_{k=2,Out:10-11PM}$ | 0.3390   | 0.2520 | $\alpha_{k=5,Out:9-10PM}$ | 0.7086   | 0.2156         |
| $\alpha_{k=3,Out:8-9PM}$ | 1.0921   | 0.2672 | $\alpha_{k=5,Out:10-11PM}$ | 1.4593   | 0.2643         |
| $\alpha_{k=3,Out:9-10PM}$ | -0.5711  | 0.1889 | $\alpha_{k=5,Out:10-11PM}$ | 1.4593   | 0.2643         |
| Viewer Segment | Parameter | Average Value | Parameter | Estimate | Standard Error | Parameter | Estimate | Standard Error |
|----------------|-----------|---------------|-----------|----------|----------------|-----------|----------|----------------|
| 1              | $\xi^\mu_{ABC}$ | 1.879         | $\xi^m_{ABC}$ | 1.926    | 1.096          | $\xi^\mu_{CBS}$ | 1.530    | 1.173    | 0.703          |
|                | $\xi^\mu_{CBS}$ | 0.587         | $\xi^m_{CBS}$ | 0.247    | 0.130          | $\xi^\mu_{FOX}$ | 0.842    | 0.318    | 0.191          |
|                | $\xi^\mu_{NBC}$ | 1.437         | $\xi^m_{NBC}$ | 1.034    | 0.502          | $\xi^\mu_{FOX}$ | 0.482    | 0.277    | 0.186          |
| 2              | $\xi^\mu_{ABC}$ | 1.118         | $\xi^m_{ABC}$ | 1.304    | 0.699          | $\xi^\mu_{CBS}$ | 0.585    | 0.318    | 0.191          |
|                | $\xi^\mu_{CBS}$ | 0.585         | $\xi^m_{CBS}$ | 0.318    | 0.191          | $\xi^\mu_{FOX}$ | 0.243    | 0.348    | 0.292          |
|                | $\xi^\mu_{NBC}$ | 0.748         | $\xi^m_{NBC}$ | 2.093    | 1.504          | $\xi^\mu_{FOX}$ | 0.482    | 0.277    | 0.186          |
| 3              | $\xi^\mu_{ABC}$ | 2.140         | $\xi^m_{ABC}$ | 10.279   | 12.060         | $\xi^\mu_{CBS}$ | 1.520    | 0.438    | 0.332          |
|                | $\xi^\mu_{CBS}$ | 1.520         | $\xi^m_{CBS}$ | 14.015   | 24.132         | $\xi^\mu_{FOX}$ | 0.482    | 0.277    | 0.186          |
|                | $\xi^\mu_{NBC}$ | 1.258         | $\xi^m_{NBC}$ | 2.805    | 1.537          | $\xi^\mu_{FOX}$ | 0.611    | 0.207    | 0.108          |
| 4              | $\xi^\mu_{ABC}$ | 1.835         | $\xi^m_{ABC}$ | 3.868    | 2.488          | $\xi^\mu_{CBS}$ | 1.371    | 0.647    | 0.315          |
|                | $\xi^\mu_{CBS}$ | 1.371         | $\xi^m_{CBS}$ | 0.647    | 0.315          | $\xi^\mu_{FOX}$ | 0.482    | 0.277    | 0.186          |
|                | $\xi^\mu_{NBC}$ | 1.457         | $\xi^m_{NBC}$ | 2.805    | 1.537          | $\xi^\mu_{FOX}$ | 0.611    | 0.207    | 0.108          |
| 5              | $\xi^\mu_{ABC}$ | 2.354         | $\xi^m_{ABC}$ | 10.750   | 14.753         | $\xi^\mu_{CBS}$ | 1.191    | 0.835    | 0.561          |
|                | $\xi^\mu_{CBS}$ | 1.191         | $\xi^m_{CBS}$ | 0.835    | 0.561          | $\xi^\mu_{FOX}$ | 0.583    | 0.034    | 0.072          |
|                | $\xi^\mu_{NBC}$ | 1.533         | $\xi^m_{NBC}$ | 3.800    | 3.707          | $\xi^\mu_{FOX}$ | 0.583    | 0.034    | 0.072          |

Precision of the advertising signal $\xi^a$ 1.172 0.635
Table 7

| $X_{j,t} = 0$ | 1 segment | 2 segments | 3 segments | 4 segments | 5 segments | 6 segments |
|---------------|-----------|------------|------------|------------|------------|------------|
| $\rho_1$      | 0.5512    | 0.4626     | 0.4433     | 0.4006     | 0.3789     | 0.3581     | 0.3605     |
| $\rho_2$      | -0.0477   | -0.0490    | -0.0482    | -0.0403    | -0.0374    | -0.0327    | -0.0362    |

$N_{i,j,t}^a$ $\rho_1 N_{i,j,t}^a + \rho_2 (N_{i,j,t}^a)^2$

| $N_{i,j,t}^a$ | $\rho_1 N_{i,j,t}^a + \rho_2 (N_{i,j,t}^a)^2$ |
|---------------|-----------------------------------------------|
| 0             | 0                                             |
| 1             | 0.5034                                        |
| 2             | 0.9114                                        |
| 3             | 1.2239                                        |
| 4             | 1.4409                                        |

Note: The models are estimated (1) under the assumption that the individual is fully informed ($\varphi_{i,j}^m = \infty$ for every $i$ and $j$), and (2) with the $\rho$ parameters the same for all segments.

Table 8

| $\zeta^a$ | Log-likelihood | Likelihood ratio |
|-----------|----------------|-----------------|
| 0.6       | -36277.594     | 7.460           |
| 0.8       | -36274.825     | 1.922           |
| 1         | -36273.875     | 0.023           |
| 1.17      | -36273.864     | ----            |
| 1.4       | -36274.261     | 0.794           |
| 1.6       | -36275.024     | 2.321           |
| 1.8       | -36276.066     | 4.405           |

Note: the values in the last column are equal to $2[\log(L(\zeta^a)) + 36273.864]$
Table 9
The relationship between the estimates of $\rho$ and $\xi^a$

(a) The estimates of $\rho_1$ and $\rho_2$ (averaged across the unobserved segments) for various values of $\xi^a$

| $\xi^a$ | $\hat{\rho}_1$ | $\hat{\rho}_2$ |
|---------|----------------|----------------|
| 0.6     | 0.443484       | -0.05077       |
| 0.8     | 0.442163       | -0.05114       |
| 1       | 0.441501       | -0.05179       |
| 1.17    | 0.440841       | -0.05173       |
| 1.4     | 0.437389       | -0.05105       |
| 1.6     | 0.436755       | -0.05138       |
| 1.8     | 0.436755       | -0.05838       |

Note: The model was estimated seven times, and in each case $\xi^a$ was set equal to a different value. The table reports the weighted average of the estimates of $\rho_1$ and $\rho_2$ across the unobserved segments.

(b) The correlation between the estimates of $\rho$ and $\xi^a$ in the model’s final solution

| Segment number | $\hat{\rho}_1$ | $\hat{\rho}_2$ |
|----------------|----------------|----------------|
| 1              | -0.03101       | 0.02118        |
| 2              | -0.03427       | 0.02541        |
| 3              | 0.02605        | -0.01217       |
| 4              | 0.03061        | -0.03765       |
| 5              | 0.00881        | -0.01885       |
| Table 10a | The effect of advertising exposures on $\theta$ | Number of exposures to advertisements |
|-----------|-----------------------------------------------|----------------------------------------|
|           | Segment 1                                      | 0   | 1   | 2   | 3   | 4   |
|           | ABC                                             | 0.4752 | 0.3654 | 0.2974 | 0.2509 | 0.2170 |
|           | CBS                                             | 0.5424 | 0.3785 | 0.2919 | 0.2379 | 0.2010 |
|           | NBC                                             | 0.5612 | 0.3817 | 0.2902 | 0.2345 | 0.1968 |
|           | FOX                                             | 0.7035 | 0.2935 | 0.1857 | 0.1358 | 0.1071 |
|           | Segment 2                                      | 0.4183 | 0.2867 | 0.2194 | 0.1781 | 0.1500 |
|           | ABC                                             | 0.6206 | 0.2750 | 0.1775 | 0.1312 | 0.1040 |
|           | CBS                                             | 0.2357 | 0.1706 | 0.1342 | 0.1107 | 0.0943 |
|           | NBC                                             | 0.4171 | 0.1394 | 0.0837 | 0.0598 | 0.0465 |
|           | FOX                                             | 0.1699 | 0.1555 | 0.1434 | 0.1331 | 0.1241 |
|           | Segment 3                                      | 0.7148 | 0.4585 | 0.3348 | 0.2647 | 0.2192 |
|           | ABC                                             | 0.0818 | 0.0761 | 0.0711 | 0.0667 | 0.0628 |
|           | CBS                                             | 0.6213 | 0.2481 | 0.1553 | 0.1130 | 0.0889 |
|           | NBC                                             | 0.3153 | 0.2626 | 0.2250 | 0.1969 | 0.1750 |
|           | FOX                                             | 0.1699 | 0.1555 | 0.1434 | 0.1331 | 0.1241 |
|           | Segment 4                                      | 0.6473 | 0.4065 | 0.2988 | 0.2369 | 0.1966 |
|           | ABC                                             | 0.3321 | 0.2618 | 0.2163 | 0.1843 | 0.1606 |
|           | CBS                                             | 0.756 | 0.3060 | 0.1935 | 0.1415 | 0.1115 |
|           | NBC                                             | 0.1777 | 0.1633 | 0.1511 | 0.1407 | 0.1315 |
|           | FOX                                             | 0.5450 | 0.3442 | 0.2541 | 0.2021 | 0.1681 |
|           | Segment 5                                      | 0.2749 | 0.2268 | 0.1932 | 0.1683 | 0.1492 |
|           | ABC                                             | 0.9092 | 0.3210 | 0.1956 | 0.1408 | 0.1099 |
|           | CBS                                             | 0.3415 | 0.2711 | 0.2270 | 0.1962 | 0.1732 |
|           | NBC                                             | 0.5978 | 0.3758 | 0.2785 | 0.2221 | 0.1851 |
|           | FOX                                             | 0.3647 | 0.2670 | 0.2128 | 0.1776 | 0.1529 |
|           | Weighted average                               | 0.7087 | 0.2791 | 0.1745 | 0.1269 | 0.0998 |
|           | Average over networks and segments              | 0.5032 | 0.2983 | 0.2232 | 0.1807 | 0.1527 |

Note: The numbers in the table are equal to the average $\theta_{i,j,t}$ across individuals and time.
### Table 10b
The effect of advertising exposures on $\sigma_{\text{ad}}$

| Segment 1 | ABC     | 0.3783 | 0.3613 | 0.3410 | 0.3222 | 0.3056 |
|-----------|---------|--------|--------|--------|--------|--------|
|           | CBS     | 0.4223 | 0.4068 | 0.3788 | 0.3532 | 0.3313 |
|           | NBC     | 0.4307 | 0.4171 | 0.3873 | 0.3601 | 0.3370 |
|           | FOX     | 0.6013 | 0.5963 | 0.5086 | 0.4478 | 0.4040 |
| Segment 2 | ABC     | 0.5092 | 0.4543 | 0.4099 | 0.3757 | 0.3486 |
|           | CBS     | 0.6572 | 0.5932 | 0.5049 | 0.4449 | 0.4017 |
|           | NBC     | 0.5288 | 0.4601 | 0.4123 | 0.3768 | 0.3491 |
|           | FOX     | 1.0048 | 0.7035 | 0.5621 | 0.4812 | 0.4273 |
| Segment 3 | ABC     | 0.2589 | 0.2497 | 0.2413 | 0.2337 | 0.2268 |
|           | CBS     | 0.3827 | 0.4274 | 0.4000 | 0.3711 | 0.3462 |
|           | NBC     | 0.2453 | 0.2372 | 0.2298 | 0.2231 | 0.2170 |
|           | FOX     | 0.7071 | 0.6250 | 0.5231 | 0.4569 | 0.4104 |
| Segment 4 | ABC     | 0.3483 | 0.3290 | 0.3116 | 0.2963 | 0.2829 |
|           | CBS     | 0.4360 | 0.4405 | 0.4065 | 0.3752 | 0.3491 |
|           | NBC     | 0.3992 | 0.3709 | 0.3463 | 0.3255 | 0.3077 |
|           | FOX     | 0.5693 | 0.5916 | 0.5064 | 0.4465 | 0.4031 |
| Segment 5 | ABC     | 0.2508 | 0.2424 | 0.2348 | 0.2278 | 0.2214 |
|           | CBS     | 0.4988 | 0.4646 | 0.4201 | 0.3842 | 0.3556 |
|           | NBC     | 0.3720 | 0.3472 | 0.3262 | 0.3083 | 0.2929 |
|           | FOX     | 0.3835 | 0.6156 | 0.5217 | 0.4569 | 0.4107 |
| Weighted average | ABC | 0.3478 | 0.3287 | 0.3102 | 0.2940 | 0.2800 |
|           | CBS     | 0.4613 | 0.4482 | 0.4091 | 0.3761 | 0.3493 |
|           | NBC     | 0.4018 | 0.3774 | 0.3511 | 0.3285 | 0.3095 |
|           | FOX     | 0.6095 | 0.6142 | 0.5183 | 0.4540 | 0.4084 |
| Average over networks and segments | 0.4551 | 0.4421 | 0.3972 | 0.3632 | 0.3368 |

Note: The numbers in the table are equal to the average $\sigma_{\text{ad}}$ across individuals and time.
Table 11  
Advertising deters consumption using the structural estimates

| $N_{i,j,t}^a$ | Among individuals who watch TV in time slot $t$, the percent of them who choose alternative $j$, that satisfies the following conditions: | Number of observations |
|---------------|-------------------------------------------------------------------------------------------------|-------------------------|
|               | $\xi_{i,j,t} < \mu_{i,j}$ | $\xi_{i,j,t} > \mu_{i,j}$ |                          |
|               | (a) All the time slots |                                      |                         |
| 0             | 43.88 | 56.12 | 6493                      |
| 1             | 31.49 | 68.51 | 99                        |
| 2 +           | 33.51 | 66.49 | 24                        |
|               | (b) The television was turned on in the previous time slot |                                      |                         |
| 0             | 45.15 | 54.85 | 5143                      |
| 1             | 32.91 | 67.09 | 91                        |
| 2 +           | 32.01 | 67.99 | 22                        |
|               | (c) All the time slots |                                      |                         |
| 0             | 40.98 | 59.02 | 3942.3                    |
| 1             | 24.19 | 75.81 | 55.76                     |
| 2 +           | 32.66 | 67.34 | 23.70                     |
|               | (d) The television was turned on in the previous time slot |                                      |                         |
| 0             | 42.04 | 57.96 | 3088.1                    |
| 1             | 26.99 | 73.01 | 49.82                     |
| 2 +           | 31.06 | 68.94 | 21.70                     |

Notes:
1. For all the observations in this table the number of advertising exposures is the same at time $t$ across all the networks.
2. The number of observations in the second panel of the table accounts for the fact that $\Pr(\xi_{i,j,t} < \mu_{i,j} - 0.5\sigma_{\xi}) + \Pr(\xi_{i,j,t} > \mu_{i,j} + 0.5\sigma_{\xi}) < 1$
| Table 12 | The matching effect of advertising |
|----------|-----------------------------------|
|          | $N^a_{i,j,t}$ | $N^a_{i,j,t} + 1$ |
| **All time slots (percent)** | |
| Chosen alternative | |
| First best | 0.9467 | 0.9556 |
| Second best | 0.0475 | 0.0408 |
| Third best | 0.0040 | 0.0030 |
| Fourth best | 0.0007 | 0.0003 |
| $U$ | 4.1462 | 4.1638 |
| t-value | 92.062 |
| **Time slots in which the individual just turned the TV on (percent)** | |
| Chosen alternative | |
| First best | 0.3203 | 0.5537 |
| Second best | 0.5650 | 0.3282 |
| Third best | 0.1068 | 0.0972 |
| Fourth best | 0.0077 | 0.0184 |
| $U$ | 1.3367 | 1.4957 |
| t-value | 183.43 |
| Variable                  | Definition                                                                                       | Actual          | Best response |          |
|--------------------------|-------------------------------------------------------------------------------------------------|-----------------|---------------|----------|
|                          |                                                                                                 | Estimate  | S.E.     | Estimate  | S.E.     |
| Demographic_Match\(j,k\) | The number of matches between the demographic characteristics of both shows. The demographic characteristics are age, gender, family status and race | 0.223     | 0.097    | 0.245     | 0.103    |
| Genre_Match\(j,k\)      | A binary variable that is equal to one if the shows are from the same genre, and zero otherwise. | 0.651     | 0.191    | 0.462     | 0.204    |
| Preceding_Show\(j,k\)   | A binary variable that is equal to one if show \(k\) directly precedes show \(j\), and zero otherwise. | 1.794     | 0.364    | 2.509     | 0.399    |
| Same_Hour\(j,k\)        | A binary variable that is equal to one if shows \(k\) and \(j\) are broadcast on the same hour, and zero otherwise. | 0.370     | 0.185    | 0.115     | 0.203    |
| Rating\(k\)             | The number of people who watch show \(k\) in our data set.                                       | -0.272   | 3.273    | 33.496    | 4.050    |
| Constant                 | ---                                                                                             | -2.019   | 0.367    | -4.627    | 0.453    |
| McFadden \(R^2\)        |                                                                                                 | 0.0692   | 0.1677   |           |          |

**Note:** The dependent variables in these logit estimations are the following. For the “actual” columns, it is a binary variable that is equal to one if an advertisement for show \(j\) appeared in show \(k\), and zero otherwise. For the “best response” columns, it is a binary variable that is equal to one if it is “optimal” (based on our model) to place an advertisement for show \(j\) in show \(k\), and zero otherwise.