Advertising, the Matchmaker[¤]

Bharat N. Anand[y]
Harvard Business School

Ron Shachar[z]
Tel-Aviv University

Preliminary. Please do not cite

Abstract

In this study, we model advertising content as an unbiased noisy signal on product attributes. Contrary to previous studies that modeled the number of advertisements that an individual is exposed to (advertising intensity) as part of the utility function, we formulate advertising content as part of the information set.Our approach yields the following implications. First, in some cases, exposure to advertising decreases the consumer's tendency to buy that product. Second, an increase in advertising intensity leads to better matching between consumers and products.We show how one can distinguish between the e¤ect of advertising on utility and on the information set using a panel dataset coupled with data on advertising exposures. Using a dataset that was designed and created to test this model and its implications, we show that this theory is supported by the data. Using the structural estimates, we show that an exposure to one advertisement decreases the consumer's probability of making a mistake by 27%..

[¤]We are grateful to Dmitri Byzalov for excellent research assistance, Steve Berry, Michael Keane, Paul Klemperer, Ariel Pakes, Peter Reiss, John Rust, Manuel Trajtenberg, Ken Wolpin, and participants in seminars at Econometrics in Tel-Aviv, Haifa, Harvard, Hebrew University, MIT, NBER Productivity Seminar, NYU, Oxford, University of Pennsylvania, Rochester, Stanford, Tel-Aviv, and Yale, for comments, and to several colleagues for helpful discussions. Anand is grateful to the Division of Research at Harvard Business School for …nancial support.

[y]Competition and Strategy Group, Soldiers Field Road, Boston, MA 02163. Phone: (617) 495-5082; Fax: (617) 495-0355; email: banand@hbs.edu.

[z]Eitan Berglas School of Economics, Tel Aviv University, Tel Aviv, Israel, 69978. Phone: 972-3-640-9202; Fax: 972-3-640-9908; 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 studies that modeled the number of advertisements that an individual is exposed to (advertising intensity) as part of the utility function, we formulate advertising content as part of the information set.Our approach yields the following implications. First, in some cases, exposure to advertising decreases the consumer's tendency to buy that product. Second, an increase in advertising intensity leads to better matching between consumers and products.We show how one can distinguish between the effect of advertising on utility and on the information set using a panel dataset coupled with data on advertising exposures. Using a dataset that was designed and created to test this model and its implications, we show that this theory is supported by the data. Using the structural estimates, we show that an exposure to one advertisement decreases the consumer's probability of making a mistake by 27%.

Grossman and Shapiro (1984) were the first to identify the role of advertising in matching consumers with products.In their setting, advertising conveys full and accurate information about the characteristics of products. Heterogeneous consumers, who have no source of information other than advertising, seek to purchase the product that best fits their needs. They conclude that market-determined levels of advertising are excessive. They also find that decreased advertising costs may reduce profits by increasing the severity of price competition. 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.The more realistic assumptions are that consumers have other sources of information other than advertisements, such as word-of-mouth, media coverage, previous experience with a good, and the profile of multiproduct firms. Furthermore, we do not assume that advertising conveys full and accurate information on attributes, but rather that it is an unbiased noisy signal. Moreover, while Grossman and Shapiro do not deal with persuasive advertising, we do allow advertising to enter the model not only through the information set, but also via a direct effect on the utility function.

It is straightforward to show that in this setting, an individual's expected utility from a product is a weighted average of the following sources of information: advertising, word-of-mouth, media coverage, previous experience, and the distribution of product attributes within each multiproduct firm . From the researcher's point of view, some of these sources are observed while others are not. Specifically, we observe the multiproduct firm profile, and the number of advertisements that an individual was exposed to, but not the other sources.

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 the match between the promoted product's attributes and consumers' tastes is below the average match of this consumer with all available products. The intuition behind this result is as follows. Without any product speci…c signals, the expected utility of the consumer is equal to her match with the average product in this market. Any product-speci…c signal that she receives shifts her expected utility towards her true utility from that product. Advertising is such a product-speci…c signal. Thus, whenever her match with a product is lower than her average match, her tendency to purchase a product diminishes with the number of advertisements that she is exposed to. In Grossman-Shapiro, 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 …rm, and thus her probability of buying such a product is zero.

Like Grossman-Shapiro, we also …nd 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 signi…cant challenges. The …rst is the distinction between the direct e¤ect of advertising on utility and its e¤ect through the information set. The identi…cation of the two e¤ects rests on the …rst implication of the model. Speci…cally, consumers di¤er in their response to advertising intensity, and that these di¤erences are correlated with their heterogeneity of preferences over product attributes. Notice that even without the e¤ect of advertising through the information set, one would expect consumer responses to advertising intensity to be heterogeneous. These heterogeneous responses are already accounted for in the standard model. However, the augmented model o¤ers another source of heterogeneity. This additional source is what identi…es the e¤ect of advertising through the information set.

The large number of unobserved variables in this model presents the second challenge. As researchers, we do not observe the prior distribution of individuals on the attributes of each product, the product-speci…c signals (through word-of-mouth, media coverage, and previous experience), and, obviously, the noisy advertising signals as they are perceived by individuals. Furthermore, we allow individuals' preferences for products and …rms to di¤er 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). 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, etc. While very useful to examine some theories of advertising, this types of data does not suit our model well for the following reasons. First, these are experience goods. Thus, signi…cant attributes of the products are not well-de…ned 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 di¢cult to track the di¤erent prices that consumers face with respect to each of the products. Prices di¤er by …rms, over time, and across consumers (through coupon schemes).

We created a dataset designed to overcome these di¢culties. 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-speci…c. Finally, almost all the commercials for television shows appear on TV. This enables us to create a comprehensive dataset of exposures to ads. Speci…cally, we were fortunate to obtain 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 …nd that, as expected, individuals' responses to advertising intensity 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 demographics of individuals with those of show cast members. We then calculate a variable that is equal to the match value from the product selected by the individual. As predicted by the model, we …nd that this variable is a positive function of advertising intensity. In these two tests, as in other non-structural examinations that are reported in section 5, we control for the persuasive nature of advertising. We do that by focusing on a sub-sample of individuals who were exposed to the same number of advertisements from all the …rms.

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 di¤er only in their exposures to advertising are identical. In other words, advertising does not have any informational role. The estimate of the precision of advertising signals is positive and statistically di¤erent from zero at the 0.1% signi…cance 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 …nd that the precision of three advertisements is equal to the precision of all other sources of information about the product together.

The availability of individual-level data on consumption and advertising exposures has generated interesting …ndings by Erdem and Keane (1996), Ackerberg (2001), and Shum (2000). 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 di¤er from their approach by allowing advertising intensity to enter the utility function; by estimating the precision versus the variance of advertising (which enables us to test the existence of advertising in the information set); and by focusing on search goods that have observable characteristics (which reveals the matching role of advertising including its ability to deter consumption). This last di¤erence expresses itself in another way. Their identi…cation 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.

Two other studies of advertising are based on individual-level panel data and discrete choice models—Ackerberg (2001) and Shum (2000). Their model of advertising is, however, very di¤erent from the one presented here. Ackerberg, following Milgrom and Roberts (1989), focuses on an experience good and models advertising intensity as a signal of product quality. Notice that unlike our approach, Ackerberg focuses on the type of advertising that is referred to in Milgrom and Roberts as "having little or no obvious informational content." Using panel data on choices of di¤erent 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 ads then were inexperienced consumers. Although Shum's (2000) model does not deal with any informational role of advertising, his …ndings are somewhat similar. In his model, habit, not experience, is the source of the di¤erent responses by consumers to advertising exposures.[1]

2 The Model

Here, we introduce the utility function, the information set, and the implications of the model.

We study di¤erentiated 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,

[1] In an early attempt to use our data (see our 1998 study), we …nd preliminary evidence for the informational role of advertising. Speci…cally, we …nd that consumers' response to advertising intensity is weaker for well-known products than for newly introduced ones.

hereafter BLP). Like BLP, products in this model are search goods.[2] 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:

$$U_{i,j} \quad = \quad X_j \beta_i + (\xi_j + \epsilon_{i,j}) + g(N^a_{i,j}) \quad \text{for } j = 1,\ldots,J \tag{1}$$

The first element of the utility represents the match between the products' observed attributes, X_j, and the preferences of the individual, β_i.[3] The parameter vector β_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 β 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, $(\xi_j + \epsilon_{i,j})$. Common effects are captured by the parameter ξ_j, while personal effects are represented by the random variable $\epsilon_{i,j}$.[4] The parameter ξ_j is often referred to as the 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 individual i is exposed to for product j; $N^a_{i,j}$. This is the modeling approach adopted by previous empirical studies.[5] It assumes that advertising can change the preferences of individuals. 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:

$$U_{i,J+1} = \gamma_i + (\xi_{J+1} + \epsilon_{i,J+1}) \tag{2}$$

[2] An individual can know her utility from a search good even without consuming it. See Tirole (1989, page 106).

[3] The variable X_j is a K-dimensional row-vector, and the parameter β_i is a column-vector of the same size.

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

[5] For example, Nevo (2001).

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

2.2 Information set

Unlike most discrete choice models, we assume that the individual is uncertain about product attributes, η_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,

$$\xi_{i,j} \equiv \eta_j + X_j\beta_i. \tag{3}$$

The prior distribution of $\xi_{i,j}$ is:[6]

$$\xi_{i,j} \sim N(\mu_i, \frac{1}{\varsigma_i^{1}}) \tag{4}$$

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 η_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 of the distribution of these attributes in the market.[7] Notice that the expectation μ_i and the precision ς_i^{1} differ across individuals because the taste parameter β_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. Each signal is independently distributed as:

$$\tilde{S}_{i,j,n}^m = \xi_{i,j} + \omega_{i,j,n}^m \text{ where } \omega_{i,j,n}^m \sim N(0, \frac{1}{\varsigma^m}) \tag{5}$$

and $n = 1,...,N_{i,j}^m$. We assume that these signals are noisy ($\frac{1}{\varsigma^m} > 0$) and unbiased. The noisiness can

[6] The normality of the prior distribution results from a normality assumption about η_j and X_j.

[7] 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.

result from various sources. For example, even previous experience is not a precise signal because of limited memory and other human information-processing mechanisms.[8]

The content of each advertisement serves the individual as a signal on product attributes. Speci…cally, each such signal is independently distributed as:

$$\tilde{S}^{a}_{i;j;n} \;=\; »_{i;j} + !^{a}_{i;j;n} \text{ where } \; !^{a}_{i;j;n} \;\; » \;\; N(0; \tfrac{1}{\&^{a}}) \tag{6}$$

and $n = 1:::N^{a}_{i;j}$.

We assume that the signals are noisy, that is $\frac{1}{\&^{a}} > 0$. 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 …nd that 29% of these were miscomprehended by consumers.[9] We assume that the signals are independent for two reasons: (1) …rms occasionally use di¤erent advertisements for the same product; (2) di¤erent exposures to the same advertisement can lead to di¤erent impressions. The independence assumption does not a¤ect our qualitative results.[10]

The e¤ect of advertisements through the information set is captured by $\&^{a}$. If $\&^{a} = 0$, then advertisements are too noisy to convey any information about product attributes. In other words, when $\&^{a} = 0$ the information sets of two individuals who di¤er only in N^{a} are the same. On the other hand, when $\&^{a} > 0$, the information sets of such consumers di¤er. Thus, $\&^{a}$ 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", $»_{i;j}$, we start by calculating her expected attribute utility.

Individual i updates her prior using the product-speci…c signals to form her expected attribute-

---

[8] Consumer learning through the miscellaneous sources has been the focus of various studies (Crawford and Shum 2000 studied dynamic learning through past experience; xxx studied the network e¤ects of word-of-mouth; and xxx studied the e¤ect of media coverage). Our focus is on advertising signals and thus these processes are degenerate in this model.

[9] They …nd similar results in their 1989 study, which uses a survey of 1,250 consumers who were exposed to print ads.

[10] The unbiasedness assumption rests on truth-in-advertising regulations. Furthermore, if a …rm 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.

utility, $\mu^p_{i;j}$:[11]

$$\mu^p_{i;j} = \frac{1}{\varsigma^p_{i;j}}\left[\varsigma^\mu_i \mu_i + \varsigma^a \sum_{n=1}^{N^a_{i;j}} S^a_{i;j;n} + \varsigma^m \sum_{n=1}^{N^m_{i;j}} S^m_{i;j;n}\right] \tag{7}$$

where $\varsigma^p_{i;j} = \varsigma^\mu_i + N^a_{i;j}\varsigma^a + N^m_{i;j}\varsigma^m$, and $S^a_{i;j;n}$ and $S^m_{i;j;n}$ are the realizations of the signals. Notice that $\frac{1}{\varsigma^p_{i;j}}$ is the variance of her posterior distribution.

Since $S^a_{i;j;n} = \xi_{i;j} + \omega^a_{i;j;n}$ where $\omega^a_{i;j;n}$ is the realization of $\tilde{\omega}^a_{i;j;n}$ and $S^m_{i;j;n} = \xi_{i;j} + \omega^m_{i;j;n}$ where $\omega^m_{i;j;n}$ is the realization of $\tilde{\omega}^m_{i;j;n}$, we can re-write equation (7) as:

$$\mu^p_{i;j} = \left[\theta_{i;j}\mu_i + (1 - \theta_{i;j})\xi_{i;j}\right] + \omega_{i;j} \tag{8}$$

where $\theta_{i;j} \equiv \frac{\varsigma^\mu_i}{\varsigma^p_{i;j}}$, and $\omega_{i;j} \equiv \frac{1}{\varsigma^p_{i;j}}\left(\varsigma^a \sum_{n=1}^{N^a_{i;j}} \omega^a_{i;j;n} + \varsigma^m \sum_{n=1}^{N^m_{i;j}} \omega^m_{i;j;n}\right)$.

>From equation (8), one can see that with a finite number of product-specific signals, $\mu^p_{i;j} \neq \xi_{i;j}$. 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 because $\theta_{i;j} > 0$ and $\omega_{i;j}$ is not equal to 0.[12]

Recall that $\mu_{i;j}$ is, by definition, 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 $N^a_{i;j}$ and ς^a. Since advertising is informative, an increase in the number or precision of advertisements would increase the informedness (thus, decreasing $\theta_{i;j}$) of the individual. The effect of $N^a_{i;j}\varsigma^a$ on $\theta_{i;j}$ is a function of ς^m, ς^μ, 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 (ς^μ is larger), (2) the other product-specific signals provide more information ($N^m_{i;j}\varsigma^m$ is larger).

The expected attribute-utility is a negative function of $N^a_{i;j}$ when $\mu_i > \xi_{i;j}$ and a positive function otherwise. As mentioned above, $\theta_{i;j}$ is a negative function of $N^a_{i;j}$. Thus, an increase in the number of advertisements decreases the weight on μ_i and increases the weight on the actual attribute-utility. Whenever $\mu_i > \xi_{i;j}$, such an increase in the number of advertisements leads to a

[11] See DeGroot [1989].

[12] The probability that $\omega_{i;j} = 0$ is equal to 0.

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. This means that advertising improves the matching between individuals and products.

The expected utility of the individual is a function of $\mu^p_{i,j}$, the persuasive effect of advertising, and $\varepsilon_{i,j}$. Specifically,

$$E\left[U_{i,j} | IS_{i,j}\right] = (\mu^p_{i,j} + \varepsilon_{i,j}) + g(N^a_{i,j}) \tag{9}$$

where $IS_{i,j}$ is the information set of individual i on product j, and $IS_{i,j} = \{\mu_i; \{S^a_{i,j,n}\}; \{S^m_{i,j,n}\}; \varsigma^1_i; \varsigma^m; \varsigma^a\}$. 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 $\xi_{i,j}$, the researcher does not observe the signals but has X_j and estimates of η_j and β_i, and thus an estimate of $\xi_{i,j}$. 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 variance $\sigma^\omega_{i,j} \equiv \frac{\varsigma^a N^a_{i,j} + \varsigma^m N^m_{i,j}}{(\varsigma^p_{i,j})^2}$. In order to write the expected utility from the researcher's point of view, we replace $\omega_{i,j}$ by $\sigma^\omega_{i,j} z_{i,j}$ where $z_{i,j}$ is a standard normal random variable. The expected utility is then:

$$u_{i,j} = \left[\theta_{i,j}\mu_i + (1 - \theta_{i,j})\xi_{i,j}\right] + \sigma^\omega_{i,j} z_{i,j} + g(N^a_{i,j}) + \varepsilon_{i,j} \text{ for } j = 1, \ldots, J \tag{10}$$

The variance $\sigma^\omega_{i,j}$ in equation (10) is another measure (in addition to $\theta_{i,j}$) of how ill-informed the individual is. When $\sigma^\omega_{i,j} = \theta_{i,j} = 0$, the expected utility is exactly equal to the utility. This would happen, for example, if the miscellaneous signals are not noisy ($\frac{1}{\varsigma^m} = 0$). In this case, the individual is fully informed. Finally, note that $\sigma^\omega_{i,j}$ is a negative function of $N^a_{i,j}$. This results from the informative nature of advertising in this model.

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

In addition to z_i, the researcher does not observe β_i, γ_i, and ς_i^m. Let

$$\begin{pmatrix} \beta_i \\ \gamma_i \\ \varsigma_i^m \end{pmatrix} = \begin{pmatrix} \beta_0 \\ \gamma_0 \\ \varsigma_0^m \end{pmatrix} + \Pi Y_i + \upsilon_i \tag{11}$$

where Y_i is a $l \times 1$ vector of demographic variables, Π is a $(K + J + 1) \times l$ matrix of coefficients that measure how the taste characteristics vary with demographics, and υ_i is a random vector whose density function is f_υ.

Let $W_{i;j}$ be the set of all variables observed by the researcher, and Ω be the set of all the parameters that are common across individuals.

Let

$$A_{i;j}(W_{i;j};\Omega) = \{(\varepsilon_i; \upsilon_i; z_i) \mid u_{i;j} \geq u_{i;r};\ \text{for } r = 1;...;J+1\}. \tag{12}$$

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:

$$p_{i;j} = \int_{A_{i;j}} dP^*(\varepsilon; \upsilon; z) \tag{13}$$

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.[13]

[13] 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 $\omega_{i;j;n}^a$ is zero, this idiosyncratic effect cancels out in $p_{i;j}$.

Since the sign of $\xi_{\iota,j} - \mu_i$ depends on β_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 we expect to find that consumer responses to advertising are positive on average, heterogeneous across consumers, and that this heterogeneity depends on β_i.

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 matching of 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 $\sum_{j=1}^{J} N^a_{i;j}$.

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 $N^a_{i;j}$. 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 increases with $N^a_{i;j}$. This correlation depends negatively on the variance of the variables which are unobserved by the researcher. In the model, this variance depends on $\sigma^{\omega}_{i;j}$. Since $\sigma^{\omega}_{i;j}$ is a negative function of $N^a_{i;j}$, we get this implication.[14]

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-

[14] 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 ε_i, the smaller the observed correlation between Y_i and X_i. A role of $\omega_{i;j}$ in our model is similar to the role of ε_i in this simple regression.

level exposures to advertisements (promoting television shows). The data on individual characteristics and choices 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) the products are differentiated, (2) consumers are heterogenous, (3) consumers are uncertain about product attributes, and (4) the researcher observes few of the product attributes. Previous studies show that data on television viewing choices satisfy these requirements.[15] 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 other products, including yogurt, ketchup, toothpaste, and coffee.

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.

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.[16] Accounting for the opportunity cost of leisure, spending on television consumption is high.

[15] 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 (2001) 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.

[16] 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 have coded the show attributes for 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 …ve show categories—for example, comedies and action dramas—and show that viewers di¤er 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, which was found by previous studies to be similar.[17]

Shows were also characterized by their cast demographics. Shachar and Emerson (2000) demonstrate that the demoraphic 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.[18] 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-speci…c. 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

[17] See Goettler and Shachar (2002).

[18] Using 1990 Census data, the sample is designed to re‡ect 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.[19] 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 (which aired on Friday at 10:00 PM). 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. 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.

Besides the importance of advertising on television channels, This makes it particularly meaningful to study the role of advertising within this context.

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

[19] 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.

structural estimation, we allow the advertising parameters to di¤er across these two parts of the week.

For the Wednesday through Friday shows, the mean number of advertisements aired per show is 4.9, and the median is 4.[20] On average, an individual is exposed to 0.37 advertisements for each show on Wednesday through Friday. A more meaningful measure of exposures to advertisements is given by conditioning on watching television in at least one time slot during Monday and Tuesday. In this case, the average exposure is xxx.

4 Preliminary evidence

In order to separate the informative e¤ect of advertising from its persuasive e¤ect, we have drawn the model's implications under the assumption that $g^0(¢) = 0$. In the empirical examination, we obviously cannot make this assumption. In the structural estimation, we simultaneously estimate the parameters of the $g(¢)$ function and $\&^a$. Here, we o¤er a non-structural approach to distinguish between the two e¤ects.

Recall that the implications predict the change in consumer behavior from an increase in the number of exposures. The di¢culty arises from the fact that such an increase alters the persuasive component of advertising. The only way to resolve this problem is by equalizing the persuasive e¤ect 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 speci…c 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. Speci…cally, 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.

Interestingly, the three shows with the highest number of advertisements were aired in the same time-slot, Thursday at 10:00 P.M, possibly indicating strategic behavior by the networks in their placement of ads.

Our non-structural approach to control for the persuasive e¤ect 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

[20] 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.

idiosyncratic reason, the three shows with the highest number of advertisements were aired in the same time slot, Thursday at 10:00 P.M.[21] The second challenge that the non-structural approach presents is the assessment of the match between consumers and products. It turns out that one of the shows at 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 news-magazines 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 ad for each network, with those who have seen 2 or more ads for each network. We …nd 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;[22] see table 2. Speci…cally, 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 news-magazines 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 …rst 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-speci…c signal. The second feature concerns …rms' 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 6.

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 signi…cant even at the 10% level. Table 3 builds on the logic of table 2. While it is 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

[21] 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 o¤er national programming after 10:00 PM.

[22] 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.

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, $Watch_{i;k}$, in the probit model is equal to one if individual i watched the newsmagazine alternative in period k, and equal to zero otherwise. Among the independent variables, there are five that serve as controls and three that examine our consumption-deterring hypothesis. These three are $N^a_{i;k}$, $NewsMatch_{i;k}$, and $Information_{i;k}$. The first, $N^a_{i;k}$, is the number of exposures of individual i to each one of the shows in time slot k. $NewsMatch_{i;k}$ 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, $Information_{i;k} = N^a_{i;k}(1 - NewsMatch_{i;k})$.[23] We expect the effect of the first two variables to be positive, and of the third to be negative. The effect of $N^a_{i;k}$ 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 $\sum_{j=1}^{J} U_{i;j} p_{i;j}$ and $\sum_{j=1}^{J} N^a_{i;j}$. 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, $Match_{i;j} = 2$.[24] Shachar and Anand (1998) show that the viewing utility is a positive function of this variable.

The test, then, is to examine the relationship between $Match^C_i \equiv \sum_{j=1}^{J} I\{C_i = j\} \cdot Match_{i;j}$ 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

[23] 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.

[24] 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.

of $Match_i^C$ as the match with the alternative that is chosen by the individual.

Once again, we conduct this test only for individuals who have been exposed to the same number of ads for all the networks, and denote this number as N_i^a. Table 3a demonstrates that $Match_i^C$ indeed increases in N_i^a. However, notice that the number of observations with positive N_i^a 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 3b-3d. In almost all 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_i^a. Consider, for example, the case comparing CBS and NBC shows for individuals who watched NBC in the previous time slot (table 3d). 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.

Advertising reduces regret A special feature of our data enables us to examine the hypothesis of informative advertising in another 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_i^a. Indeed, the correlation between the switching decisions and N_i^a is equal to -0.45. To calculate this correlation, we consider 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 4 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 $ViewingTime_i$ 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 ad. This e¤ect is signi…cant at the 0.1% level.[25]

5 Estimation and Identi…cation Issues

This section consists of …ve subsections. The …rst (4.1) extends the model presented in section 2, to account for the multiperiod nature of the data. The speci…c functional forms of the utility and the density functions of the unobserved variables are presented in 4.2. The next subsection (4.3) constructs the likelihood function, and our simulation approach is described in 4.4. The …nal subsection (4.5) discusses the identi…cation of the model's parameters.

5.1 The Multiperiod Model

In any given week, television networks o¤er multiple shows, with each show being o¤ered only once. Recall that in the simple model, each …rm o¤ers 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. However, as pointed above, each …rm (television network) has multiple products and the consumer chooses among the alternatives several times.

5.1.1 Utility

Previous studies of television viewing choices …nd strong evidence of state dependence.[26] 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 …nding. The networks tend to schedule similar shows in sequential time slots.[27] 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, 2002). While there is not a good behavioral explanation for state dependence in television viewing, we allow the utility to include a state dependence variable in order not to misspecify the model. Furthermore, following Heckman (1981), 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-speci…c. Speci…cally, the utility of individual i from alternative j in period t (where any combination of j and t de…nes a show) is:

[25] Moreover, the negative e¤ect of $N^a_{i,t}$ on the switching probability is stable across various model speci…cations, 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.

[26] See Rust and Alpert (1984).

[27] 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 …nd deviations from this strategy.

$$
\begin{aligned}
U_{i,j,t} &= X_{j,t}\beta_i + (\eta_{j,t} + \varepsilon_{i,j,t}) + g(N^a_{i,j,t}) \\
&+ \delta_{i,j,t} I\{C_{i,t-1} = j\} + \alpha_{i,j}
\end{aligned}
\tag{14}
$$

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 4.2. 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:

$$
U_{i,J+1,t} = \gamma_{i,t} + (\eta_{J+1,t} + \varepsilon_{i,J+1,t}) + \delta_{i,out,t} I\{C_{i,t-1} = J+1\} + \alpha_{i,J+1} \tag{15}
$$

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:

$$
\xi_{i,j,t} \sim N(\mu_{i,j}, \frac{1}{\varsigma_{i,j}}) \tag{16}
$$

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, 2001). 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."[28]

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, Erdem and Keane 1996, and

[28] 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."

Crawford and Shum 2000). 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:

$$
\begin{aligned}
u_{i,j,t} = {} & \left[\theta_{i,j,t}\mu_{i,j} + (1-\theta_{i,j,t})\xi_{i,j,t}\right] + \sigma^{\omega}_{i,j,t} z_{i,j,t} + g(N^{a}_{i,j,t}) + \varepsilon_{i,j,t} \\
& + \delta_{i,j,t} I\{C_{i,t-1} = j\} + \alpha_{i,j} \quad \text{for } j = 1, \ldots, J
\end{aligned}
\tag{17}
$$

where

$$
\theta_{i,j,t} \equiv \frac{\varsigma^{\mu}_{i,j}}{\varsigma^{\mu}_{i,j} + N^{a}_{i,j,t}\varsigma^{a} + N^{m}_{i,j,t}\varsigma^{m}} \text{ and } \sigma^{\omega}_{i,j,t} \equiv \left(\frac{\varsigma^{a}N^{a}_{i,j,t} + \varsigma^{m}N^{m}_{i,j,t}}{\varsigma^{\mu}_{i,j} + N^{a}_{i,j,t}\varsigma^{a} + N^{m}_{i,j,t}\varsigma^{m}}\right)^{2}
$$

The most significant difference between the expected utility of the extended model in (17) and the simple model in (10) is that μ_i is replaced by $\mu_{i,j}$, and ς^{μ}_{i} is replaced by $\varsigma^{\mu}_{i,j}$. 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 $\varsigma^{\mu}_{i,j}$. The larger is $\varsigma^{\mu}_{i,j}$, the weaker is the consumer's response. The logic goes 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 $\varsigma^{\mu}_{i,j}$ 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, ς^{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, υ_i, whose density function is f_{υ}, includes all these parameters alongside those presented earlier in equation (11).

Attribute utility We restrict Π and f_ν in equation (11) so that the match element in the utility is:

$$
\begin{aligned}
X_{j,t}\beta_i = {} & \beta_{Gender}\, I\{\text{the gender of } i \text{ and the cast of show } j,t \text{ is the same}\} \\
& + \beta_{Age0}\, I\{\text{the age group of } i \text{ and the cast of show } j,t \text{ is the same}\} \\
& + \beta_{Age1}\, I\{\text{the distance between the age group of } i \text{ and the cast of show } j,t \text{ is one}\} \\
& + \beta_{Age2}\, I\{\text{the distance between the age group of } i \text{ and the cast of show } j,t \text{ is two}\} \\
& + \beta_{Family}\, I\{i \text{ lives with her family and show } j,t \text{ is about family matters}\} \\
& + \beta_{Race}\, Income_i\, I\{\text{one of the main characters in show } j,t \text{ is African American}\} \\
& + Sitcom_{j,t}(\beta_{Sitcom} Y_i^{\beta} + \nu_i^{Sitcom}) \\
& + ActionDrama_{j,t}(\beta_{AD} Y_i^{\beta} + \nu_i^{AD}) \\
& + RomanticDrama_{j,t}(\beta_{RD} Y_i^{\beta} + \nu_i^{RD})
\end{aligned}
$$

The first six β 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, we 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.[29] The taste for different show categories is a function of observed (Y_i^{β}) and unobserved (ν_i^{Sitcom}, ν_i^{AD}, ν_i^{RD}) individual characteristics. The observed variables included in Y_i^{β} are $Teens_i$, $GenerationX_i$, $BabyBoomer_i$, $Older_i$, $Female_i$, $Income_i$, $Education_i$, $Family_i$, and $Urban_i$. As mentioned above, 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 β_{AD}^{Female}. All the other parameters are denoted accordingly.

Recall that the attribute utility is a function of the match element and ξ_t. While we can identify an $\xi_{j,t}$ 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 both have one ξ parameter. Given our intent in uncovering fundamental attributes of the shows, this is a natural restriction.

[29] 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.

State dependence parameters The utility presented in equation (14) includes a switching cost element, $\delta_{i,j,t}\mathbf{1}\{C_{i,t-1}=j\}$. 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:

$$
\begin{aligned}
&\delta_{i,j,t}\mathbf{1}\{C_{i,t-1}=j\} \\
&+\mathbf{1}\{C_{i,t-1}\neq j\}\mathbf{1}\{\text{The show on } j \text{ started at least 15 minutes ago}\}\delta_{InProgress} \\
&\text{where}
\end{aligned}
$$

$$
\delta_{i,j,t}=\begin{bmatrix}
Y_i^{\delta}\delta^{Y}+X_{j,t}^{\delta}\delta^{X}+\upsilon_i^{\delta} \\
+\delta_{First15}\mathbf{1}\{\text{The show on } j \text{ started within the past 15 minutes}\} \\
+\delta_{Last15}\mathbf{1}\{\text{The show on } j \text{ is at least one hour long and will end within 15 minutes}\} \\
+\delta_{Continuation}\mathbf{1}\{\text{The show on } j \text{ started at least 15 minutes ago}\}
\end{bmatrix}
$$

where the observed variables included in Y_i^{δ} are $Teens_i$, $GenerationX_i$, $BabyBoomer_i$, $Older_i$, $Female_i$; $Family_i$, 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 $Sport_{j,t}$.[30]

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 υ_i^{δ}.

These parameters can differ across show types as well. For example, we may expect δ 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 depend on time. Specifically, we expect δ to be small in the first 15 minutes of a show, since the viewers have not had enough time to get hooked by the show. For the same reason, we expect the state dependence to 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 been on the air for less than 15 minutes, $\delta_{InProgress}$ is expected to be negative.

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

[30] 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).

$$\delta_{i,out,t} \mathbb{1}\{C_{i,t-1} = (J+1)\} = \begin{bmatrix} Y_i^{\delta}\delta^{Y} + \upsilon_i^{\delta} \\ +\delta_{Hour} \mathbb{1}\{\text{The time is either 9:00 PM or 10:00 PM}\} \end{bmatrix} \mathbb{1}\{C_{i,t-1} = (J+1)\}$$
$$+\delta_{FOX10:00} \mathbb{1}\{C_{i,t-1} = FOX\} \mathbb{1}\{\text{The time is 10:00 PM}\}$$

Individual characteristics ($Y_i^{\delta}\delta^{Y} + \upsilon_i^{\delta}$) 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 δ to be lower at that time ($\delta_{Hour} < 0$). 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 $\delta_{FOX10:00}$ to be positive.

Outside alternative We restrict Π and f_{υ} 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^{γ} includes the variables $Basic_i$, $Premium_i$, All_i, and $Same_{i,t}$. 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 All_i (and $Same_{i,t}$) 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 (All_i, and $Same_{i,t}$) and the personal unobserved parameters $\alpha_{i,J+1}$ are designed to capture other sources of such differences. Specifically, we allow $\alpha_{i,J+1}$ to differ across the different hours of the night. Since our data set starts on Monday, the variables All_i, and $Same_{i,t}$ have missing values for this day. Thus, we include specific parameters to account for this: $\gamma_{Monday8:00}$ is added to the outside utility for the first hour on Monday night prime time, and $\gamma_{Monday9:00}$ and $\gamma_{Monday10:00}$ 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:

$$\eta_{J+1,t} = \eta_{J+1,t+12} = \eta_{J+1,t+24} = \eta_{J+1,t+36} = \eta_{J+1,t+48} \text{ for } t = 1, ..., 12.$$

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:

$$
\begin{aligned}
g(N^a_{i,j,t}) = & \left[\rho_{1,MT,i} MondayTuesday_{j,t} + \rho_{1,WF,i} WednesdayFriday_{j,t}\right] N^a_{i,j,t} \\
& + \left[\rho_{2,MT,i} MondayTuesday_{j,t} + \rho_{2,WF,i} WednesdayFriday_{j,t}\right] \left(N^a_{i,j,t}\right)^2
\end{aligned}
$$

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 effects of $N^a_{i,j,t}$ 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 $\sigma^{\mu}_{i,j}$, 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 to depends only on the distribution of the attributes of the shows which are broadcast during that part.[31]

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

[31] That is, for example, for shows aired between 8:00 to 10:00 PM, $\mu^{8-10}_{i,j} = \frac{1}{40} \sum_{t \in t^{8-10}} \xi_{i,j,t}$, where t^{8-10} is the set of all the time slots between 8:00 and 10:00 PM during the week.

choice probability is multinomial logit.

The density function f_{υ} is assumed to be discrete.[32] Specifically, $\upsilon_i = \upsilon_k$ 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 αs. Thus, we normalize $\alpha_{k,ABC} = 0$ for each type k.

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

Similarly, since all the age categories are included in Y_i^{δ}. we need to normalize the υ_i^{δ} for one of the types. To do so, we set $\upsilon_k^{\delta} = 0$ for $k = 1$. Finally, since all the show categories are included in $X_{j,t}^{\delta}$, one of the parameters in δ^X needs to be normalized. We set $\delta_{NewsMagazine} = 0$.

Since all the age categories are included in Y_i^{γ}, 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 γ parameters of the age categories by 1, the likelihood does not change. Thus, we set η 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 β_{Age} parameters, and the β_{Sitcom}, β_{AD} and β_{RD} for one age group. We do so by setting $\beta_{Age2} = \beta_{Sitcom}^{Teens} = \beta_{AD}^{Teens} = \beta_{RD}^{Teens} = 0$.

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

5.3 The likelihood function

The conditional choice probability is:

$$f_1(C_{i,t}|C_{i,t-1}, W_{i,t}, \upsilon_i, z_{i,t}, \Omega) = \frac{\sum_{j=1}^{J+1}\left[I\{C_{i,t} = j\}\exp(\bar{u}_{i,j,t}(C_{i,t-1}, W_{i,j,t}, \upsilon_i, z_{i,j,t}, \Omega))\right]}{\sum_{j=1}^{J+1}\exp(\bar{u}_{i,j,t}(C_{i,t-1}, W_{i,j,t}, \upsilon_i, z_{i,j,t}, \Omega))} \tag{18}$$

[32] The unobserved individual-specific parameters were introduced in sections 2 and 4. Together, the vector υ_i is $(\alpha_{i,ABC}, \alpha_{i,CBS}, \alpha_{i,NBC}, \alpha_{i,FOX}, \alpha_{i,OUT,8:00}, \alpha_{i,OUT,9:00}, \alpha_{i,OUT,10:00}, \upsilon_i^{Sitcom}, \upsilon_i^{AD}, \upsilon_i^{RD}, \upsilon_i^{\delta}, \rho_{1,MT,i}, \rho_{2,MT,i}, \rho_{1,WF,i}, \rho_{2,WF,i}, \varsigma_{i,ABC}^{m}, \varsigma_{i,CBS}^{m}, \varsigma_{i,NBC}^{m}, \varsigma_{i,FOX}^{m})$.

where $W_{i,j,t}$ is a vector of all the variables in the model (that is $X_{j,t}$, Y_i, and $N^a_{i,j,t}$), $W_{i,t}$ 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 $z_{i,j,t}$, Ω is the vector of the parameters that are common to all the individuals,[33] and $\bar{u}_{i,j,t} = u_{i,j,t} - \varepsilon_{i,j,t}$.

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.[34] The history probability is then:

$$f_2(C_i|W_i, \upsilon_i, z_i, \Omega) = \prod_{d=1}^{5} \left[f_1(C_{i,(12d-11)}|W_{i,(12d-11)}, \upsilon_i, z_{i,(12d-11)}, \Omega) \prod_{t=(12d-10)}^{12d} f_1(C_{i,t}|C_{i,t-1}, W_{i,t}, \upsilon_i, z_{i,t}, \Omega) \right] \tag{19}$$

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:

$$\int_{\tilde{z}_1} f_2(C_i|W_i, \upsilon_i, (\tilde{z}_1, \ldots, \tilde{z}_{64}), \Omega)\phi(\tilde{z}_1)d\tilde{z}_1$$

where $\phi(\tilde{z}_1)$ is the standard normal density function. Repeating this integration for the other 63 shows in the week gives us $f_3(C_i|W_i, \upsilon_i, \Omega)$. Recall that for any individual, $z_{i,j,t}$ is constant across all time slots of a specific show. In practise, 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

[33] That is, $\Omega = \{\eta_{j,t}, \beta_{Gender}, \beta_{Age0}, \beta_{Age1}, \beta_{Age2}, \beta_{Family}, \beta_{Race}, \beta_{Sitcom}, \beta_{AD}, \beta_{RD}, \delta^Y, \delta^X, \delta_{First15}, \delta_{Last15}, \delta_{Continuation}, \delta_{InProgress}, \delta_{Hour}, \delta_{FOX10:00}, \gamma, \varsigma^a\}$.

[34] Notice that the choice at 10:45 PM on the previous night is not the lagged choice for the 8:00 PM time slot.

unobserved z's. Indeed, the largest number of integrals for each time slot is xxx. 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, π_i, we get the marginal probability:

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

where Ω' includes Ω, the π_k's, and the λ's.

The likelihood function is:

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

5.4 Simulating the marginal probability

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

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

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

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

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

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 $\hat{\Omega}'_{MSL}$.[35] In our case this is

[35] 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 χ^2 with degrees of freedom equal to

achieved at $R = 400$.xxx

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 Berry, Levinsohn and Pakes (1994). We draw the z's from a multivariate normal approximation of each person's posterior distribution of z, given some preliminary MSL estimate of Ω^0, 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 xxx the size of the RMSE when not using importance sampling.[36]

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 $\frac{1}{\varsigma_{i;j}^{\omega}} = 0$ for all i and j). This discussion illustrates which parameters can be identified without the restrictions and additional variables of the model without this assumption.

5.5.1 Utility parameters

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

---

the number of parameters estimated.

[36] The RMSE of $\hat{f}_4(C_i|W_i;\Omega^0)$ is computed using N_R sets of R draws as:

$$RMSE(R) = \left[\frac{1}{N_R}\sum_{n=1}^{N_R}\frac{(\hat{f}_{4;n}(C_i|W_i;\Omega) - f_{4;true})^2}{f_{4;true}}\right]^{0.5} \qquad (25)$$

where $f_{4;true}$ 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 Ω^0. Quantifying the magnitude of this reduction is of interest. To our knowledge, constructing the empirical distribution of $\hat{\Omega}^0_{MSL}$ via a bootstrapping method is the only way to proceed. Unfortunately, the cpu time required to compute $\hat{\Omega}^0_{MSL}$ prohibits us from pursuing this goal.

shows do not …t her preferences well.[37]

The conditional correlation between the number of advertising exposures and viewing choices identi…es the persuasive parameters ½. The targeting strategies of …rms pose a challenge for the researcher in obtaining an unbiased estimate of ½. Recall that table 4 provided an example of such targeting strategies. The problem is that exposures to advertisements are not random. This problem is obviously not speci…c 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, di¤erences 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 re‡ect the persuasive power of advertising. Individual-level data and proper modeling of consumer preferences can resolve this problem.[38] 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-speci…c parameters. Thus, the researcher can separate the e¤ect of advertising and the role of consumer preferences. Even this approach cannot resolve the problem if …rms can target advertisements to consumers individually. However, advertising on television does not provide such opportunities to …rms.[39]

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 identi…cation of the signals' parameters.

Prior Distribution Recall that the estimation of the prior distribution parameters was already discussed in sub-section 5.2.5. Once the parameters ¯ and ´ are identi…ed, we also have all the variables which are a function of them, i.e., $»_{i;j;t}$, $¹_{i;j}$ and $\&^{1}_{i;j}$.

Since one of the advantages of our model is the empirical distinction between $®_{i;j}$ and $¹_{i;j}$, it is worth clarifying the identifying source of this distinction. In the model we set $¹_{i;j} = \frac{1}{T} \sum_{t=1} \hat{»}_{i;j;t} =$

[37] As discussed in the literature, there are various sources of identifying ± separate from ®. 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 in‡uenced by ® (and show characteristics), but not by ±.

[38] Such an approach is taken by Ackerberg (2000), Erdem and Keane (1996), and Shum (2000).

[39] This means that the larger is the audience that is exposed a commercial for a product, the more precise is the estimate of ½. This is because large audiences are likely to be more heterogeneous in their preferences. The television networks occasionally target very large audiences for reasons that are not speci…c to the weekly episode being advertised. For example, as mentioned earlier, the three shows with the largest number of advertisements were all in the same time slot, Thursday at 10:00 PM.

$\frac{1}{T}\sum_{t=1}^{T}\hat{\ }_{j,t} + \left(\frac{1}{T}\sum_{t=1}^{T} X_{j,t}\right)\hat{\ }_i$. Thus, the identification of $\mu_{i,j}$ is based on the introduction of a new explanatory variable—the mean offering of each network (for example, $\frac{1}{T}\sum_t X_{j,t}$ 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 within 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: $L_{i,j} = a_j + b \cdot (\frac{1}{T}\sum_t X_{j,t}) + \varepsilon^L_{i,j}$. The estimates of the parameters a_j and b can be used to calculate the predicted value of the "error" term, $\varepsilon^L_{i,j}$. One can now separate between the observed individual–network match, $\hat{b} \cdot (\frac{1}{T}\sum_t X_{j,t})$, and the unobserved match, $\varepsilon^L_{i,j}$. In our structural estimation, this separation between the observed match, $\mu_{i,j}$, 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 ς^m and ς^a. The sum of the precision of all the product-specific signals (advertising and miscellaneous), $\varsigma^a N^a_{i,j,t} + \varsigma^m_{i,j}$, enters $\bar{u}_{i,j,t}$ (and thus the likelihood) only through $\theta_{i,j,t}$ and $\sigma^\omega_{i,j}$. Furthermore, neither ς^a nor ς^m enters the likelihood in any other form. Thus, we start by presenting the identification of $\varsigma^a N^a_{i,j,t} + \varsigma^m_{i,j}$, and then show how one can separately identify ς^a and $\varsigma^m_{i,j}$.

The dependence of $\theta_{i,j,t}$ and $\sigma^\omega_{i,j}$ on $\varsigma^a N^a_{i,j,t} + \varsigma^m_{i,j}$ lead to three identifying factors. The first is the effect of $\mu_{i,j}$ through $\theta_{i,j,t}$ on product choices. If $\frac{1}{\varsigma^a N^a_{i,j,t} + \varsigma^m_{i,j}} = 0$, then $\theta_{i,j,t} = 0$, and the choice of a product is not a function of $(\mu_{i,j} - \xi_{i,j,t})$.[40] Thus, if in the data the difference between $\mu_{i,j}$ and $\xi_{i,j,t}$ affects the choices, $\frac{1}{\varsigma^a N^a_{i,j,t} + \varsigma^m_{i,j}} > 0$. The larger the effect, the smaller is $\varsigma^a N^a_{i,j,t} + \varsigma^m_{i,j}$. Notice that this source of identification relies on observed product and firm attributes.

The other two identifying factors concern $\sigma^\omega_{i,j}$. Recall that $\sigma^\omega_{i,j}$ is a negative function of $\varsigma^a N^a_{i,j,t} + \varsigma^m_{i,j}$. The variance $\sigma^\omega_{i,j}$ affects the observed consumer behavior in two ways. First, the larger is $\sigma^\omega_{i,j}$, the smaller is the correlation between any observed variables and choices.[41] Second, the larger is $\sigma^\omega_{i,j}$, the stronger is the unexplained persistence in choices within a show.

[40] Recall that one can rewrite equation (17) as:

$$u_{i,j,t} = \xi_{i,j,t} + \theta_{i,j,t}(\mu_{i,j} - \xi_{i,j,t}) + \sigma^\omega_{i,j,t} z_{i,j,t} + g(N^a_{i,j,t}) + \varepsilon_{i,j,t} + \delta_{i,j,t} I\{C_{i,t-1} = j\} + \alpha_{i,j} \quad \text{for } j = 1, \ldots, J \tag{26}$$

[41] 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 ε_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 ε_i in this simple regression.

Finally, $\&^a$ can be separately identi…ed from $\&^m_{i;j}$ using data on an individual's exposure to ads, $N^a_{i;j;t}$. A positive $\&^a$ leads to a negative correlation between $N^a_{i;j;t}$ and our two measures of how ill-informed the individual is, $µ_{i;j;t}$ and $¾^!_{i;j}$. Thus, if $\&^a > 0$, we would expect to …nd that an increase in $N^a_{i;j;t}$ would: (1) reduce the e¤ect of $(¹_{i;j}$ ¡ $»_{i;j;t})$ on choices; (2) increase the correlation between any observed variables and choices; and (3) decrease the unexplained persistence in choices. Recall that the …rst of these identifying factors re‡ects the implications presented in the model (that is, consumption-deterring and matchmaking). Speci…cally, since the expected utility is a positive function of $(¹_{i;j}$ ¡ $»_{i;j;t})$, an increase in $N^a_{i;j;t}$ reduces the consumers' tendency to purchase a product, when $(¹_{i;j}$ ¡ $»_{i;j;t}) > 0$ (and increases her tendency when $(¹_{i;j}$ ¡ $»_{i;j;t}) < 0$).

Each of these identifying factors is di¤erent from that of the persuasive e¤ect. As mentioned above, $½ > 0$ implies that an increase in $N^a_{i;j;t}$ should have a positive e¤ect on consumer i's tendency to purchase alternative j at time t. In contrast, $\&^a > 0$ implies that an increase in $N^a_{i;j;t}$ should have such a positive e¤ect when $(¹_{i;j}$ ¡ $»_{i;j;t}) < 0$ but negative e¤ect otherwise. This di¤erence by itself enables us to identify both ½ and $\&^a$. The dependence of $¾^!_{i;j}$ on $N^a_{i;j;t}$ further assists us in this task.

6 Results

We …rst present the estimates of the utility parameters, followed by those of the parameters in the information set, including $\&^a$.

The integral in equation (??) is evaluated numerically using importance sampling with 400 points from a Sobol sequence, as detailed in Section 4.3. The (asymptotic) standard errors are derived from the inverse of the simulated information matrix.[42]

We report the results for a model with 5 segments ($K = 5$), in tables 5(a)-5(g). 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.

6.1 Utility parameters

6.1.1 Show attributes (¯s and ´s)

A meaningful way to illustrate the e¤ects of all the parameters is in terms of probabilities. To do so, we de…ne a baseline viewer—she is 30 years old, lives alone, in an urban area, and has median income and education level. Furthermore, she is part of the largest unobserved segment.

[42] The reported standard errors, therefore, neglect any additional variance due to simulation error in the numerical integration.

Viewers prefer shows whose cast demographic is similar to their own (see table 5(a)). The age of the cast has the largest effect. For example, the probability of the baseline viewer to watch a show whose cast demographic (Generation X) matches hers is about three times larger than that of a viewer who is 55 years old but similar in other respects. We also find that viewers living with their family like to watch shows about families; viewers prefer to watch shows with a cast of the same gender as their own, and low income people are more likely to watch shows whose cast is African-American.

Viewers differ in both observed and unobserved ways in their taste for particular show types. Generation X viewers like psychological dramas more than action dramas and sitcoms. Coupled with the results above, this might explain why the networks choose a Generation X cast for such shows. Older viewer and “baby boomers” prefer action dramas to psychological dramas or sitcoms. And, women prefer psychological dramas over the other show categories. This finding is consistent with common beliefs about the viewing habits of women.People living in urban areas tend to like action dramas less than those living in rural areas. Finally, we find that sitcoms also tend to be preferred by viewers of higher income, and psychological dramas preferred by low-income and less-educated viewers.

After controlling for differences in choices based on observed characteristics of individuals and shows, there are large differences in the “unexplained popularity” of shows; see table 5(b). The show with the highest “unexplained popularity” is the sitcom Home Improvement (ABC), and the one with the lowest $\eta_{j,t}$ is the generation X romantic drama Beverly Hills. Notice that Beverly Hills was aired twice during this week. Its $\eta_{j,t}$ in the regular time is high, while the “unexplained popularity” of the irregular time (Monday at 9:00) is the lowest among all shows. To illustrate these differences, the conditional probability that the baseline individual would watch Home Improvement is about 10 times bigger than the probability that she would watch Beverly Hills.

6.1.2 State dependence (δ)

Our findings are consistent with those of previous studies documenting state dependence in television viewing choices. We find that state dependence is the most important source for observed network loyalty within a night. For example, our model predicts that the probability of watching a sitcom conditional on the viewer having watched that sitcom in the previous time slot is 81%.[43] The state dependence parameters (δs) are presented in Table 5(c). The predicted persistence for sitcom above is based on the average across viewer types ($\delta = 1.47$) and the additional state dependence specific to sitcoms ($\delta_{sitcom} = 0.7826$). The analogous conditional probabilities for watching an action drama is 76%, for a romantic drama is 81%, whereas for sports shows is 53%, and for

[43] This assumes that $U_{i,out,t} \approx 2.5$; and $U_{i,j,t} = 0$ for any network j.

news-magazines, 66%. State dependence appears to be higher for shows where there is a plot line of some kind that can hook viewers, as is the case with sitcoms and romantic dramas.

State dependence varies across viewer types. The conditional probabilities above (of watching a sitcom on any network, having watched it in the previous time slot) range from 72% to 88% across viewer types. Similarly, for viewers with access to cable channels, hence more program o¤erings on TV, state dependence is lower (the conditional probability above for watching a sitcom on a network channel reduces to 76% for viewers with cable).State dependence in choices is, as is commonly thought, higher for female viewers compared to males, although, somewhat surprisingly, there is not much di¤erence across age groups.

Additional parameters help explain when viewers get hooked to a show. We …nd that the persistence is highest in the last …fteen minutes of a drama, and the lowest in the …rst …fteen minutes. Since many non-network shows end on the hour one should expect to …nd lower state dependence for the outside alternative at that time. Indeed we …nd that viewers are most likely to tune in to network TV at the beginning of each hour ($\pm_{hour} = ¡0{:}273$). And, since FOX programming switches to a¢liates at 10 PM (therefore FOX viewers included in outside option after that), we treat FOX viewers at 9:45 separately ($\pm_{FOX10:00} = 0{:}764$). Finally, we …nd that, as expected, switching into the middle of a show without watching its beginning is costly ($\pm_{InProgress} = ¡0{:}49$).

6.1.3 Preference for the outside option (°)

Viewers with cable access tend to watch less network TV; see table 5(d) for this and other ° parameters. The unconditional probability of watching any network is 5%, compared to 6% for those with no cable access.[44]

There appear to be clear patterns in the times at which viewers tune in to watch network TV. First, some viewers can simply be categorized as network TV lovers ($°_{all} = ¡0{:}576$)—the probability of some viewers tuning in to network TV during any time slot of the week is higher than for others. Second, viewers tend to watch at particular times in the evening—the "same time slot" e¤ect is large ($°_{same} = ¡0{:}672$), and markedly signi…cant.[45]

There are clear di¤erences in the preference for network TV across age groups. Young viewers watch it the least, older viewers the most. This may re‡ect the existence of popular cable channels for young viewers, like MTV and ESPN, and the fact that young people have a higher utility from non television activities. Females, and viewers with higher incomes watch less TV. But, there is

[44] On the other hand, the fact that this di¤erence is not large may re‡ect a demand-driven preference for cable access as well: those who don't have cable do not watch much TV at all, hence the probability of watching network TV is lower for these viewers.

[45] Since the variables $All_{i;j;t}$ and $Same_{i;j;t}$ have missing values on Monday (the beginning of the sample), it is not surprising that we …nd an additional negative e¤ect for this day ($°_{Mon\_8}$; $°_{Mon\_9}$ and $°_{Mon\_10}$ are all negative).

not much di¤erence across education, family, and urban groups. While this may seem surprising at …rst, recall that we only analyze prime time TV viewing habits here: it is likely that there might be signi…cant di¤erences across such viewer demographic groups in their tendency to watch daytime programs. We also …nd that the utility from the outside alternative increases during the evening.[46]

6.1.4 Individual-brand match (®)

After conditioning on observed individual and show characteristics, and controlling for …rm attributes and state dependence, viewers display unobserved loyalty towards particular networks; see table 5(e). The largest segment of viewers are more likely to watch ABC and NBC than the other networks.[47] 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. These results suggest that there may be important unobserved characteristics in the "image" of networks that appeal to particular groups of viewers.

6.1.5 Persuasive E¤ect

Advertising has a signi…cant direct positive e¤ect on utility. While the persuasive e¤ect di¤ers signi…cantly across viewer segments (see table 5(f)), all segments are persuaded by advertising. As mentioned earlier, we estimated a set of parameters for shows on Monday and Tuesday, and another set for those on Wednesday through Friday because we have missing data on advertisements for shows in the earlier part of the week. It turns out, however, that the two sets of parameters are quite similar, therefore we restrict our discussion to the set of parameters for shows only on Wednesday through Friday.

There are two sensible ways to assess the e¤ectiveness of advertising. The …rst is by studying the change in the probability of watching a show when exposed to a single ad for it. The second is by examining the peak probability of viewing as the viewer is exposed to varying numbers of advertisements. It turns out that, when using either of these measures, the persuasive e¤ect of advertising is the strongest for the …fth segment—exposure to a single ad more than doubles the viewing probability for such individuals (going from 6.18% to 14.27%). The "wear-out" aspect of the persuasive e¤ect is also evident after exposure to more than six ads. The …rst six ads increases the viewing probability from 6.2% to 60%. In contrast, the largest viewer segment (segment 1) is

[46] 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 °$_{8:00}$ estimate. This parameter is biased, since we are missing the lagged choice of 7:45. Most viewers (about 75% of them) have their television o¤ at 7:45, and thus they have a large switching cost to turn on the television. Since we do not include the lagged choice from 7:45, our estimate of °$_{8:00}$ is upward biased. Notice, though, that this bias does not a¤ect the parameters of interest in this study.

[47] This is based on the average of the shows' "unexplained popularity".

least persuaded by advertising. While the first ad increases the probability of watching a show by 33% (from 6.2% to 8.2%), the second ad increases the viewing probability only by an additional 15%, and the effect of the third ad is virtually zero.

6.2 Information Set Parameters (ζ)

As described in the model, individuals obtain information on product attributes from three sources: (1) the distribution of product attributes within each multiproduct firm, (2) miscellaneous product-specific signals, and (3) advertising. Here we describe the estimated precision of each one of these information sources in order.

The importance of a firm's product line in providing information about product attributes ($\zeta^1_{i;j}$) depends (negatively) on the diversity of product attribute-utilities ($\xi_{i;j;t}$) for each network. Specifically, since our estimate of the utility-attribute is $\hat{\xi}_{i;j;t} = \alpha_{j;t} + X_{j;t}\hat{\beta}_i$, it follows that $\hat{\zeta}^1_{i;j} = \left[\frac{1}{T-1}\sum_t\left(\hat{\xi}_{i;j;t} - \frac{1}{T}\sum_t \hat{\xi}_{i;j;t}\right)^2\right]^{-1}$. Averaging $\hat{\zeta}^1_{i;j}$ across individuals in each viewer segment, we find that, for all viewer segments, ABC is the most informative "brand" and FOX is the least clear (the average of $\hat{\zeta}^1_{i;j}$ across viewers types is: $\hat{\zeta}^1_{i;ABC} = 1.82$; $\hat{\zeta}^1_{i;CBS} = 1.17$, $\hat{\zeta}^1_{i;NBC} = 1.28$ and $\hat{\zeta}^1_{i;FOX} = 0.53$).The finding for FOX may seem surprising, since, as discussed earlier, this network appears to offer the most homogeneous profile of shows: many GenerationX dramas, and no sitcoms or news-magazines.However, recall that the attribute-utility is a function of both vertical attributes $\alpha_{j;t}$ and horizontal ones, $X_{j;t}$. While the variation in $X_{FOX;t}$ is indeed the lowest among the four networks, the variance of its vertical attributes ($\alpha_{j;t}$) is the highest. In other words, while viewers may know what type of show and cast demographic to expect when they tune into FOX, they are much less certain about show "quality".

Recall that the weight the individual puts on "firm-level information" is a positive function of its clarity. Thus, the less diverse the product attributes of a firm, the larger the effect of $\mu_{i;j}$ on choices. In reality, of course, not "all things are equal". An individual relies on $\mu_{i;j}$ when she is uncertain about the attributes of a product. Conversely, when she is well informed about such attributes—because of word-of-mouth, previous experience, or advertising—she should place less weight on such information. Next, we discuss the miscellaneous sources of information and examine their effect on $\theta_{i;j;t}$.

The parameter $\zeta^m_{i;j}$ represents the precision of the miscellaneous signals. Thus, a high $\zeta^m_{i;j}$ indicates that viewer i is very familiar with the shows of network j: We find that on average viewers are familiar with ABC's and NBC's shows more than with the shows of the other networks (average $\zeta^m_{i;ABC} = 4.95$; average $\zeta^m_{i;NBC} = 3.58$). In contrast, viewers are much less familiar with shows on CBS and FOX (average $\zeta^m_{i;CBS} = 0.81$, $\zeta^m_{i;ABC} = 0.21$). These findings pass a "reality check" quite easily. To see this, note that one would expect that $\zeta^m_{i;j}$ would be a function of the ratings of the shows and their “age” (i.e., the number of seasons that they were on the air before 1995). There are two reasons to …nd a positive relationship between ratings and $³^m_{i;j}$: (1) successful shows have people chatting about them, thus creating “word of mouth” information, and (2) viewers are more likely to have previously watched such popular 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 NBC’s highest rated shows in 1995 were in their …rst 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 $³^m_{i;j}$ 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.

The di¤erences across viewer segments in the precision of information obtained from “miscellaneous” sources is clear; see table 5(f). It varies from $\hat{³}^m_{i;FOX} = 0{:}034$ for the …fth segment to $\hat{³}^m_{i;NBC} = 14$ for the third segment. The third segment is the best informed about television shows, and is also well informed about ABC’s shows ($\hat{³}^m_{i;ABC} = 10$). It is worth pointing out that the people of this segment also like watching network television shows the most—their outside utility is the smallest compared with the other types. These results illustrate the di¤erences across viewers in their prior information about shows.

Table 6 demonstrates the e¤ect of …rm-level information and the miscellaneous sources of information on the weights µ. The …rst column in the table presents the values of µ for each segment and network for the case that $N^a_{i;j;t} = 0$ for every i, j, and t, using the estimates of $³^m_{i;j}$ and $³^1_{i;j}$, and the fact that the relative weight placed $^1{}_{i;j}$ by viewers who have not been exposed to any ads is given by $µ_{i;j} = \frac{³^1_{i;j}}{³^1_{i;j} + ³^m_{i;j}}$. Although ABC and NBC are the networks with the “clearest” image, viewers place the lowest weight on them since they are well-informed about their shows. In contrast, FOX is more important in the role of network image in guiding viewers’ choices. The highest µ is $0{:}942$ for the …fth viewer segment with respect to FOX—as mentioned above, this type knows very little about FOX’s shows. On the other hand, the lowest µ is 0.082 for the third type with respect to NBC—as mentioned above, this segment knows NBC’s shows quite well.

6.2.1 Precision of advertising signals

One can think of the parameters discussed so far as controls for the key parameter of interest, $\&^a$. The model presented in this study suggests a new role for advertising—its e¤ect through the consumer’s information set. This approach has several novel implications. The test of this approach rests on whether $\&^a > 0$ or not. We …nd that the data support the theory—the estimate of $\&^a$ is di¤erent from zero, and statistically signi…cant. The informative e¤ect of advertising also turns out to be behaviorally important. Below, we outline the speci…c results.

Our estimate of 3^a is 1.172 with a standard error of 0.635. Notice that since 3^a cannot be negative, a more appropriate way to assess its signi…cance level is based on its e¤ect on the likelihood. We have estimated the model for various (…xed) values of 3^a. As one might expect, the relationship between the log-likelihood and 3 is not symmetric. We can reject the hypothesis that $3^a = 0$ at the 0.001 signi…cance level. It is also worth noting that we have estimated various versions of the model, and in each of these the estimate of 3^a was positive and had similar e¤ects on choices. Thus, the results of the structural estimation clearly establish that individuals use the information in the content of the ads when making choices.

The informative e¤ect of advertising is not just statistically signi…cant, but behaviorally important as well. Since the average $3^m_{i;j}$ across viewers and networks is $2{:}4$; this implies that, on average, exposure to two ads provides the same amount of information as all other miscellaneous sources of information that viewers get about shows. Similarly, the average $3^1_{i;j}$ across viewers and networks is 1.2, implying that the information about a show contained in a single advertisement is equivalent to the information obtained from the information embodied in a …rm's product line choice ($1_{i;j}$).

The e¤ect of advertising exposures on µ is quite dramatic, as illustrated in table 6. Each row in the table presents our calculation of the average µ if the number of ads that each viewer is exposed to for each show is N. If there was no advertising, viewers would have placed the same weight on …rm-level information ($1_{i;j}$) as on the product attributes when making choices. With exposure to a single ad, the weight that an average viewer places on $1_{i;j}$ drops from 0.5 to 0.381. As expected, the e¤ect of the …rst exposure is the most dramatic—as seen in the table, ads are much more e¤ective for viewers who have not seen any ads than for those who have seen at least one ad, because the former are less informed about product attributes. One can also study how the informative e¤ect of advertising varies with the prior information of viewers, by comparing the e¤ect of ads on µ across di¤erent viewer segments and across di¤erent networks. For example, for the …fth type, μ_{FOX} drops from 0.943 to 0.571 with exposure to a single ad, while for the third viewer segment, the weight on ABC's "brand image' hardly changes with exposure to ads.

Another way to assess the informative value of advertising is through its e¤ect on the uncertainty faced by individuals about product attributes, as measured by the variance of (the posterior on) product attributes. On average, the variance drops from 0.33 for viewers who have not seen any advertisements, to about 0.1 for those who have seen four ads.

Other meaningful ways to assess the informative e¤ect of advertising is by: illustrating how exposure to ads a¤ects choices, and (2) demonstrating its e¤ect on matching. This is discussed in section 7.

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

Consumer choices were simulated under two scenarios, which differ in the number of advertisements that consumers were exposed to. In the first scenario, we use the actual data on advertising exposures, and in the second, each consumer is exposed to an additional advertisement for each show. Focusing on consumers who chose to watch TV, we find that while the percent of viewers who make the "right" choice under the first scenario is 75.5 percent, it is 81.2 percent under the second scenario. Notice that since the researcher is not uncertain about product attributes, one can calculate the alternative that yields the highest utility for the individual. This alternative is termed the "right" choice above. Another measure of the extent to which advertisements improve the matching of consumers and products is the increase in utility per time slot that is experienced by the viewer from her choice under the two scenarios. This number is 0.05, and is statistically different from zero at the 0.001% level. Since $\gamma_{\text{Basic}} = 0.2$ per time slot, the increase in utility as a result of an additional exposure to one advertisement equals 25% of the increase in utility from having a cable connection.[48]

We are currently evaluating the consumption-deterring role of advertising using our estimates.

7.2 Targeting Strategies

We are currently characterizing the networks' targeting strategies, calculating the profit-maximizing targeting strategies based on our model, and comparing the two. Furthermore, we are evaluating the bias in estimates of advertising effects that results from misspecification of the information set.

[48] These results are obtained using 100 simulation draws (of the product-specific signals and $\varepsilon_{i,j,t}$) for each individual, and under the following assumptions: all the δ's are equal to zero, and the persuasive effect does not change as a result of the increase in exposures. The rationale behind these assumptions is the following. Using the estimates of δ leads to two difficulties in evaluating the results: first, the percent of correct choices is even higher than the ones reported in the text; and, second, since the consumer is assumed to be myopic, the state dependence effect might distort the measure of the matchmaking role of advertising.

8 Table 1

Variables de…ning individual characteristics

(all the binary variables are equal to either zero or one)

$Teens_i$	a binary variable that is equal to one if the individual's age is between 6 and 17
$GenerationX_i$	a binary variable that is equal to one if the individual's age is between 18 and 34
$BabyBoomer_i$	a binary variable that is equal to one if the individual's age is between 35 and 49
$Older_i$	a binary variable that is equal to one if the individual's age is 50 and over
$Income_i$	On unit interval, 0 = less than $10,000, 1 = $40,000 and over
$Education_i$	On unit interval, 0 = 0-8 years grade school, 1 = 4 or more years college
$Basic_i$	a binary variable that is equal to one if the individual has a basic cable service
$Premium_i$	a binary variable that is equal to one if the individual has basic and premium service
$Female_i$	a binary variable that is equal to one if the individual is a woman
$Family_i$	a binary variable that is equal to one if the individual lives with his or her family
$Urban_i$	a binary variable that is equal to one if the individual lives in an urban area

Variables de…ning show characteristics

(all the following are binary variables that are equal to one if the condition holds, and to zero otherwise)

$ActionDrama_{j;t}$	the show of network j on period t is an action drama
$PsychologicalDrama_{j;t}$	the show of network j on period t is a psychological drama
$Sitcom_{jt}$	the show of network j on period t is a situation comedy
$Sport_{jt}$	the show of network j on period t is a sport event
$MondayTuesday_{j;t}$	the show of network j on period t is on Monday or Tuesday
$WednesdayFriday_{j;t}$	the show of network j on period t is on Wednesday, Thursday or Friday

References

[1] Abernethy, Avery and George Franke (1996) "The Information Content of Advertising: A Meta-Analysis," Journal of Advertising, 25 (Summer), 1-17.

[2] Ackerberg, D. (1998). "Advertising, Learning, and Consumer Choice in Experience Good Markets: A Structural Empirical Examination." Mimeo.

[3] Akerlof, George and Rachel Kranton (2000) "Economics and Identity," Quartely Journal of Economics, 115 (3), 715-754.

[4] Anand, Bharat and Ron Shachar (2000a) "Multiproduct Firms, Information, and Loyalty ", Harvard Business School Working Paper No. 00-069.

[5] Anand, Bharat and Ron Shachar (2000b) "Informative Advertising can be Persuasive," Mimeo.

[6] Anderson, Simon and Stephen Coate, "Market Provision of Public Goods: The Case of Broadcasting," Working Paper.

[7] Becker G. and K. Murphy (1993) "A Simple Theory of Advertising as a Good or Bad," Quarterly Journal of Economics, 108, 941-964.

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

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

[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, Z., D. Horksy, and Y. Raban (1988). "An Empirical Dynamic Model of Optimal Brand Choice". Foerder Institute of Economic Research, Working Paper no. 88.

[15] Erdem, T. (1998). "An Empirical Analysis of Umbrella Branding", Journal of Marketing Research, 35, 339-351.

[16] Erdem, T. and M. Keane (1996) "Decision-Making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets," Marketing Science, 15 (1), 1-20.

[17] Goettler, Ron and Ron Shachar (2000) "An Empirical Analysis of Spatial Competition," The Foerder Institute for Economic Research Working Paper No. 12-2000.

[18] Grossman, G. and C. Shapiro (1984) "Informative Advertising with Di¤erentiated 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] Jacoby, Jacob and Wayne Hoyer (1982) "Viewer Miscomprehension of Televised Communication: Selected Findings," Journal of Marketing, 46 (4), 12-26.

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

[23] Mankiw, Gregory (2000) Principles of Economics, Harcourt College Publishers.

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

[25] McFadden, Daniel (1973), "Conditional Login Analysis of Qualitative Choice Behavior," in Frontiers in Econometrics, ed. by P. Zarembka, New York: Academic Press.

[26] Milgrom P. and J. Roberts (1986) "Price and Advertising Signals of Product Quality," Journal of Political Economy, 94, 796-821.

[27] Moshkin, Nickolay and Ron Shachar (2000) "Switching Costs or Search Costs?" The Foerder Institute for Economic Research Working Paper No. 3-2000.

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

[29] Niederreiter, H. (1978), "Quasi-Monte Carlo Methods and Pseudo-Random Numbers," Bulletin of the American Mathematical Society, 84, 957–1028.

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

[31] Papageorgiou, A. and Traub, J.F. (1996), "New Results on Deterministic Pricing of Financial Derivatives," Technical Report CUCS-028-96, Department of Computer Science, Columbia University.

[32] Resnk, Alan and Bruce Stern (1977) "An Analysis of Information Content in Television Advertising," Journal of Marketing, 41 (1), 50-53.

[33] Roberts, M.J. and L. Samuelson (1988). "An Empirical Analysis of Dynamic, Nonprice Competition in an Oligopolistic Industry," Rand Journal of Economics, 19 (2), 200-220.

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

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

[36] Schmalensee, R. (1972). The Economics of Advertising. Amsterdam: North-Holland.

[37] 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.

[38] Shachar, R. and B. Anand (1998) "The E¤ectiveness and Targeting of Television Advertising." Journal of Economics and Management Strategy, 7 (3), 363-396.

[39] Shachar, Ron (1994) "A Diagnostic Test for the Sources of Persistence in Individuals' Decisions," Economics Letters, 45, 7-13.

[40] Shapiro, Carl and Hal Varian (1998). Information Rules: A Strategic Guide to the Network Economy. Harvard Business School Press.

[41] Shum, Matthew (1999) "Advertising and Switching Behavior in the Breakfast Cereals market," Mimeo.

[42] Stegeman, Mark (1991) "Advertising in Competitive Markets," American Economic Review, 81 (1), 210-223.

[43] Telser, L. (1964) "Advertising and Competition," Journal of Political Economy, 72, 537-562.

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

TABLE I.

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. Summary Statistics: Individual Demographic Characteristics		
Variable	Mean	Standard Deviation
Teens	0.0627	0.2425
Generation-X	0.2400	0.4272
Baby Boom	0.2764	0.4474
Old	0.4191	0.4936
Female	0.5319	0.4991
Male	0.4681	0.4991
Family	0.4304	0.4953
Income	0.8333	0.2259
Education	0.7421	0.2216
Urban	0.4149	0.4929
Basic	0.3642	0.4813
Premium	0.3588	0.4798

Notes: Definition of variables are provided in the Appendix.

Table 2. Advertising deters consumption: Effect of Ad Exposures on Propensity to View *48 Hours*		
Exposures to ads 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. First number in each cell of the table indicates percentage of viewers within that cell who watched 48 hours. Second number in each cell denotes total number of individuals within that cell.

2. "Low" Match indicates that the viewer was tuned in to newsmagazine shows no more than 10% of the time that these shows were aired during the rest of the week (i.e., Dateline NBC on Tuesday 10 p.m. and Wednesday 9 p.m. Prime Time Live on Wednesday 10 p.m. 20/20 on Friday 10 p.m. and Dateline NBC on Friday 9 p.m.). "High" Match indicates that the viewer was tuned in to newsmagazines shows at least than 10% of the time watching these shows.

Table 3a. Advertising and Welfare: Effect of Ad Exposures on Chosen "Match" between Viewer and Show	
Exposures to Ads on ABC, CBS, and NBC	Did not watch CBS or NBC in previous time slot
0	0.3679 (0.1757) 3581
1 or more	0.3914 (0.1517) 177

Notes:

1. First number in each cell denotes average "chosen match" for viewers in that cell, where "chosen match" is defined on page 21 in the text. Second number in each cell denotes standard deviation of "chosen match" for viewers in that cell. Third number in each cell denotes total number of individuals in that cell.

Table 3b. Advertising and Welfare: Effect of Ad Exposures on Chosen "Match" between Viewer and Show			
Exposures to Ads on ABC and CBS	Did not watch ABC or CBS in previous time slot	Watched ABC in previous time slot	Watched CBS in previous time slot
0	0.5317 (0.2028) 1184	0.4751 (0.1976) 1271	0.5691 (0.1977) 773
1	0.5120 (0.1820) 72	0.5529 (0.1930) 225	0.5477 (0.1815) 108
2 or more	0.5544 (0.1687) 13	0.5959 (0.1585) 60	0.5929 (0.1826) 23

Table 3c. Advertising and Welfare: Effect of Ad Exposures on Chosen "Match" between Viewer and Show			
Exposures to Ads on ABC and NBC	Did not watch ABC or NBC in previous time slot	Watched ABC in previous time slot	Watched NBC in previous time slot
0	0.5385 (0.2031) 1291	0.5732 (0.2021) 1316	0.5247 (0.1903) 840
1	0.5177 (0.2037) 90	0.5031 (0.1901) 169	0.5666 (0.1896) 258
2 or more	0.5856 (0.1467) 15	0.5021 (0.1431) 32	0.5843 (0.1717) 85

Table 3d. Advertising and Welfare:
Effect of Ad Exposures on Chosen "Match" between Viewer and Show

Exposures to Ads on CBS and NBC	Did not watch CBS or NBC in previous time slot	Watched CBS in previous time slot	Watched NBC in previous time slot
0	0.5341 (0.2106) 1103	0.5824 (0.2104) 678	0.4967 (0.1953) 850
1	0.5268 (0.1857) 84	0.5213 (0.1997) 123	0.5462 (0.2020) 214
2 or more	0.5735 (0.2032) 28	0.5478 (0.1699) 61	0.5938 (0.1723) 113

Table 4. Advertising and Repeat Purchase: Probit estimates of effect of advertising exposures on propensity to switch from a show		
Variable	Coefficient	Standard Error
Advertising Exposure	-0.0217	0.0054
Viewing Time	-0.0095	0.0071
Constant	0.2562	0.0169
N	2665	

Notes:

1. For each individual i and time slot t, $Viewing\ Time_{i,t}$ measures the fraction of time an individual spent watching TV during the other nights of the week.

Table 5(a). Structural Estimates: Preferences for Show Attributes					
Parameter	Coefficient	Standard Error	Parameter	Estimate	Standard Error
β_{Gender}	0.2207	0.0438	β_{PD}^{Teens}	0.0000	----
β_{Age0}	1.1038	0.0956	β_{PD}^{GenX}	-0.2543	0.2482
β_{Age1}	0.7977	0.0753	$\beta_{PD}^{BabyBoomer}$	-0.3948	0.2363
β_{Age2}	0.0000	----	β_{PD}^{Older}	-0.4102	0.2506
β_{Family}	0.3628	0.1125	β_{PD}^{Female}	0.7535	0.1256
β_{Race}	-0.8338	0.2513	β_{PD}^{Income}	-1.4195	0.2696
β_{Sitcom}^{Teens}	0.0000	----	$\beta_{PD}^{Education}$	-0.6431	0.2800
β_{Sitcom}^{GenX}	-0.7701	0.1800	β_{PD}^{Family}	0.1810	0.1328
$\beta_{Sitcom}^{BabyBoomer}$	-0.8261	0.1781	β_{PD}^{Family}	0.0556	0.1146
β_{Sitcom}^{Older}	-1.1824	0.1926	$\beta_{Sitcom}^{k=1}$	0.0000	----
β_{Sitcom}^{Female}	0.3756	0.0818	$\beta_{AD}^{k=1}$	0.0000	----
β_{Sitcom}^{Income}	-0.2850	0.2146	$\beta_{PD}^{k=1}$	0.0000	----
$\beta_{Sitcom}^{Education}$	-0.2578	0.2068	$\beta_{Sitcom}^{k=2}$	0.8258	0.2451
β_{Sitcom}^{Family}	0.1857	0.1075	$\beta_{AD}^{k=2}$	-0.4012	0.2110
β_{Sitcom}^{Urban}	-0.0490	0.0819	$\beta_{PD}^{k=2}$	-2.2048	0.4688
β_{AD}^{Teens}	0.0000	----	$\beta_{Sitcom}^{k=3}$	0.4971	0.2532
β_{AD}^{GenX}	-0.4530	0.1873	$\beta_{AD}^{k=3}$	1.3646	0.2300
$\beta_{AD}^{BabyBoomer}$	-0.3628	0.1832	$\beta_{PD}^{k=3}$	0.4368	0.3361
β_{AD}^{Older}	-0.1960	0.1902	$\beta_{Sitcom}^{k=4}$	-0.5281	0.2106
β_{AD}^{Female}	0.4009	0.0848	$\beta_{AD}^{k=4}$	-0.2880	0.1959
β_{AD}^{Income}	-0.6648	0.2321	$\beta_{PD}^{k=4}$	-0.0913	0.2649
$\beta_{AD}^{Education}$	-0.0526	0.2074	$\beta_{Sitcom}^{k=5}$	-1.0587	0.3402
β_{AD}^{Family}	0.0094	0.1011	$\beta_{AD}^{k=5}$	0.0329	0.2973
β_{AD}^{Urban}	-0.1525	0.0879	$\beta_{PD}^{k=5}$	-0.1440	0.4272

Table 5(b). Structural Estimates: Unobserved "Show Quality" Parameters

Network	Show	Estimate	Standard Error	Network	Show	Estimate	Standard Error
ABC	*The Marshal*	-2.8790	0.4857	CBS	*48 Hours*	-0.7520	0.4892
ABC	*Monday Night Football*	-0.7858	0.4660	CBS	*Here Comes The Bride*	-0.3655	0.4363
ABC	*Roseanne*	0.2503	0.4037	CBS	*Ice Wars*	-0.3727	0.3895
ABC	*Hudson Street*	-0.3880	0.4233	NBC	*Fresh Prince*	-0.1552	0.5400
ABC	*Home Improvement*	1.5279	0.3979	NBC	*In The House*	-0.4798	0.5681
ABC	*Coach*	-0.2780	0.3897	NBC	*She Fought Alone*	-0.9509	0.4271
ABC	*NYPD Blue*	-0.0980	0.3493	NBC	*Wings*	0.0105	0.4008
ABC	*Ellen*	-0.3270	0.4061	NBC	*NewsRadio*	-0.1312	0.4305
ABC	*The Drew Carey Show*	-0.4180	0.4187	NBC	*Frasier*	0.2670	0.3785
ABC	*Grace Under Fire*	0.2358	0.3656	NBC	*Pursuit of Happiness*	-1.3123	0.4257
ABC	*The Naked Truth*	-0.4921	0.3947	NBC	*Dateline NBC (T)*	-0.2772	0.4537
ABC	*Prime Time Live*	-0.3662	0.4420	NBC	*Seaquest 2032*	-1.3784	0.3738
ABC	*Columbo*	-0.5516	0.3223	NBC	*Dateline NBC (W)*	-0.3429	0.3962
ABC	*Murder One*	-1.7448	0.3765	NBC	*Law and Order*	-0.8912	0.3449
ABC	*Family Matters*	1.2251	0.4465	NBC	*Friends*	0.7674	0.3888
ABC	*Boy Meets World*	0.2548	0.4357	NBC	*The Single Guy*	0.1014	0.3967
ABC	*Step by Step*	0.0264	0.3797	NBC	*Seinfeld*	0.5041	0.3680
ABC	*Hangin' With Mr. Cooper*	0.8526	0.4200	NBC	*Caroline in the City*	-0.2557	0.3803
ABC	*20/20*	-0.0328	0.4308	NBC	*E.R.*	0.7088	0.3145
CBS	*The Nanny*	-0.9715	0.6057	NBC	*Unsolved Mysteries*	-0.2550	0.3550
CBS	*Can't Hurry Love*	-1.4918	0.6663	NBC	*Dateline NBC (F)*	-0.7417	0.4040
CBS	*Murphy Brown*	-2.4282	0.6220	NBC	*Homicide*	-1.4604	0.3640
CBS	*High Society*	-2.6447	0.6959	FOX	*Melrose Place*	-1.2749	0.8488
CBS	*Chicago Hope*	-0.2969	0.6263	FOX	*Beverly Hills 90210 (M)*	-3.3298	0.7978
CBS	*The Client*	-0.8840	0.3882	FOX	*Bram Stoker's Dracula*	-1.2685	0.4060
CBS	*Nothing Lasts Forever*	0.6917	0.3907	FOX	*Beverly Hills 90210 (W)*	1.3548	0.3856
CBS	*Bless This House*	0.0083	0.4187	FOX	*Party of Five*	-1.1786	0.4072
CBS	*Dave's World*	0.7447	0.4443	FOX	*Living Single*	0.2984	0.4739
CBS	*Central Park West*	-0.8127	0.3986	FOX	*The Crew*	0.6113	0.5035
CBS	*Courthouse*	-1.3956	0.4507	FOX	*New York Undercover*	-0.3484	0.4713
CBS	*Murder, She Wrote*	-0.1609	0.3365	FOX	*Strange Luck*	-1.0986	0.4136
CBS	*New York News*	-1.4036	0.3757	FOX	*X-Files*	0.0000	----

Table 5(c). Structural Estimates: Switching Cost Parameters		
Parameter	Estimate	Standard Error
δ_{Sitcom}	0.7826	0.0959
$\delta_{ActionDrama}$	0.5266	0.0995
$\delta_{PsychDrama}$	0.7392	0.1031
δ_{News}	0.0000	----
δ_{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
δ_{Basic}	-0.3147	0.0438
$\delta_{\Pr emium}$	-0.2139	0.0463
δ_{Female}	0.1157	0.0359
δ_{Family}	-0.0223	0.0456
δ_{Teens}	0.0000	----
$\delta_{GenerationX}$	-0.0344	0.0712
$\delta_{BabryBoomer}$	-0.0486	0.0643
δ_{Older}	-0.0171	0.0737
$\delta_{Continuation}$	1.9578	0.0949
δ_{Out}	1.1590	0.1051
$\delta_{First15}$	-0.5493	0.0782
δ_{Last15}	0.4852	0.1281
δ_{Hour}	-0.2732	0.0868
$\delta_{FOX\,10:00}$	0.7644	0.1539
$\delta_{In\Pr ogress}$	-0.4920	0.0722

Table 5(d). Structural Estimates: Preference for Outside Alternatives

Parameter	Estimate	Standard Error
γ_{Basic}	0.2033	0.0387
$\gamma_{Premium}$	0.2860	0.0409
γ_{All}	-0.5756	0.0985
γ_{Same}	-0.6717	0.0615
γ_{Teens}	3.1675	0.3410
$\gamma_{GenerationX}$	2.6041	0.3663
$\gamma_{BabyBoomer}$	2.4314	0.3693
γ_{Older}	1.9424	0.3720
γ_{Female}	0.1338	0.0637
γ_{Income}	-0.4199	0.1645
$\gamma_{Education}$	-0.0925	0.1504
γ_{Family}	0.1248	0.0794
γ_{Urban}	-0.0289	0.0619
$\gamma_{Monday8:00}$	-1.3008	0.2024
$\gamma_{Monday9:00}$	-1.5725	0.2065
$\gamma_{Monday10:00}$	-1.1502	0.1648
$\gamma_{8:00}$	0.0000	----
$\gamma_{8:15}$	-1.1088	0.0972
$\gamma_{8:30}$	-0.7968	0.0993
$\gamma_{8:45}$	-0.9490	0.1302
$\gamma_{9:00}$	-0.9607	0.1366
$\gamma_{9:15}$	-0.9166	0.1571
$\gamma_{9:30}$	-0.8597	0.1517
$\gamma_{9:45}$	-0.7489	0.1666
$\gamma_{10:00}$	-0.5399	0.2649
$\gamma_{10:15}$	-0.4485	0.2780
$\gamma_{10:30}$	-0.0856	0.2770
$\gamma_{10:45}$	-0.5399	0.2649

Parameter	Estimate	Standard Error
$\alpha_{k=1,Out,8-9PM}$	0.0000	----
$\alpha_{k=1,Out,9-10PM}$	0.0000	----
$\alpha_{k=1,Out,10-11PM}$	0.0000	----
$\alpha_{k=2,Out,8-9PM}$	-0.2169	0.2562
$\alpha_{k=2,Out,9-10PM}$	0.3043	0.1643
$\alpha_{k=2,Out,10-11PM}$	0.3390	0.2520
$\alpha_{k=3,Out,8-9PM}$	1.0921	0.2672
$\alpha_{k=3,Out,9-10PM}$	-0.5711	0.1889
$\alpha_{k=3,Out,10-11PM}$	-1.8062	0.2240
$\alpha_{k=4,Out,8-9PM}$	-0.0908	0.2051
$\alpha_{k=4,Out,9-10PM}$	-0.1366	0.1361
$\alpha_{k=4,Out,10-11PM}$	-0.3435	0.1843
$\alpha_{k=5,Out,8-9PM}$	0.0123	0.3154
$\alpha_{k=5,Out,9-10PM}$	0.7086	0.2156
$\alpha_{k=5,Out,10-11PM}$	1.4593	0.2643

Table 5(e). Structural Estimates: Individual-Brand Match Parameters			
Viewer Segment	Parameter	Estimate	Standard Error
1	α_{ABC}	0	----
	α_{CBS}	0	----
	α_{NBC}	0	----
	α_{FOX}	0	----
2	α_{ABC}	0.1000	----
	α_{CBS}	0.3666	0.1545
	α_{NBC}	-0.0148	0.1357
	α_{FOX}	-0.5551	0.3225
3	α_{ABC}	0.1000	----
	α_{CBS}	0.4676	0.1801
	α_{NBC}	0.3865	0.1455
	α_{FOX}	0.2172	0.2488
4	α_{ABC}	0.1000	----
	α_{CBS}	-0.0222	0.1323
	α_{NBC}	0.3008	0.1173
	α_{FOX}	0.1474	0.1817
5	α_{ABC}	0.1000	----
	α_{CBS}	-0.0960	0.2238
	α_{NBC}	-0.0744	0.2029
	α_{FOX}	-0.3102	0.2984

Table 5(f). Structural Estimates: Information and Persuasion Parameters						
Viewer Segment	Information Parameters			Persuasion Parameters		
	Parameter	Estimate	Standard Error	Parameter	Estimate	Standard Error
1	ς^m_{ABC}	1.926	1.096	$\rho_{1,MT}$	0.3097	0.1308
	ς^m_{CBS}	1.173	0.703	$\rho_{2,MT}$	-0.0836	0.0403
	ς^m_{NBC}	1.034	0.502	$\rho_{1,WF}$	0.3865	0.0909
	ς^m_{FOX}	0.247	0.130	$\rho_{2,WF}$	-0.0777	0.0271
2	ς^m_{ABC}	1.304	0.699	$\rho_{1,MT}$	0.1159	0.2461
	ς^m_{CBS}	0.318	0.191	$\rho_{2,MT}$	0.0163	0.0749
	ς^m_{NBC}	2.093	1.504	$\rho_{1,WF}$	0.1722	0.1541
	ς^m_{FOX}	0.348	0.292	$\rho_{2,WF}$	0.0044	0.0453
3	ς^m_{ABC}	10.279	12.060	$\rho_{1,MT}$	0.4767	0.2366
	ς^m_{CBS}	0.438	0.332	$\rho_{2,MT}$	-0.0542	0.0697
	ς^m_{NBC}	14.015	24.132	$\rho_{1,WF}$	0.346	0.1598
	ς^m_{FOX}	0.277	0.186	$\rho_{2,WF}$	-0.0504	0.0464
4	ς^m_{ABC}	3.868	2.488	$\rho_{1,MT}$	0.4717	0.2013
	ς^m_{CBS}	0.647	0.315	$\rho_{2,MT}$	-0.1124	0.0716
	ς^m_{NBC}	2.805	1.537	$\rho_{1,WF}$	0.2183	0.1223
	ς^m_{FOX}	0.207	0.108	$\rho_{2,WF}$	-0.0140	0.0385
5	ς^m_{ABC}	10.750	14.753	$\rho_{1,MT}$	1.2772	0.2695
	ς^m_{CBS}	0.835	0.561	$\rho_{2,MT}$	-0.1882	0.0794
	ς^m_{NBC}	3.800	3.707	$\rho_{1,WF}$	1.0071	0.1460
	ς^m_{FOX}	0.034	0.072	$\rho_{2,WF}$	-0.0801	0.0326
Advertising Precision						
	ς^a	1.172	0.635			

Table 6. Effect of Advertising Exposures on θ

		Number of exposures to advertisements, N[a]				
Viewer	Network	0	1	2	3	4
Segment 1	ABC	0.4754	0.4091	0.3681	0.3401	0.3198
	CBS	0.5425	0.4435	0.3914	0.3590	0.3369
	NBC	0.5621	0.4534	0.3984	0.3649	0.3423
	FOX	0.7004	0.4550	0.3907	0.3610	0.3439
Segment 2	ABC	0.4187	0.3391	0.2986	0.2738	0.2570
	CBS	0.6317	0.4181	0.3590	0.3312	0.3148
	NBC	0.2355	0.1963	0.1744	0.1603	0.1505
	FOX	0.4093	0.2463	0.2134	0.1992	0.1913
Segment 3	ABC	0.1700	0.1613	0.1539	0.1477	0.1423
	CBS	0.7475	0.5736	0.4991	0.4570	0.4298
	NBC	0.0818	0.0783	0.0753	0.0727	0.0703
	FOX	0.6293	0.4005	0.3445	0.3192	0.3047
Segment 4	ABC	0.3154	0.2836	0.2609	0.2440	0.2308
	CBS	0.6499	0.5036	0.4386	0.4015	0.3773
	NBC	0.3319	0.2895	0.2620	0.2428	0.2286
	FOX	0.7432	0.4807	0.4129	0.3817	0.3638
Segment 5	ABC	0.1778	0.1691	0.1617	0.1554	0.1499
	CBS	0.5442	0.4233	0.3692	0.3382	0.3178
	NBC	0.2748	0.2458	0.2255	0.2105	0.1990
	FOX	0.9428	0.5709	0.4945	0.4613	0.4428