Quantitative Marketing and Economics, 2, 283–320, 2004.

The Risk Reduction Role of Advertising*

DMITRI BYZALOV
Department of Economics, Harvard University, Cambridge, MA 02138, USA
E-mail: byzalov@fas.harvard.edu

RON SHACHAR
Faculty of Management, Tel Aviv University, Tel Aviv, Israel 69978
E-mail: rroonn@post.tau.ac.il

Abstract. This study shows theoretically and empirically that exposure to advertising increases consumers' tendency to purchase the promoted product *because* the informative content of advertising resolves some of the uncertainty that the risk averse consumers face and thus reduces the risk associated with the product. We call this effect the "risk-reduction" role of advertising. The risk-reduction model implies that advertising effectiveness depends on (a) the risk preference parameter, (b) the precision of the advertising message, (c) the familiarity of the consumer with the product, (d) the consumer's sensitivity to products' attributes (and thus, her involvement level with the product), and (e) the diversity of products offered by multiproduct firms. These findings suggest that ads spending should be higher (a) for new and relatively unknown products, (b) for high-involvement products, (c) when ads can be quite precise, and (d) when the firm offers a diverse product-line. It also implies that ads should target consumers (a) who are more sensitive to risk, (b) who are more involved, and (c) those who are not familiar with the promoted product.

The model allows ads to affect choices also through a direct effect on the utility (i.e., the standard approach to formulate the effect of advertising). In our empirical example (where the products are television shows) the risk-reduction effect is significant and strong and the direct effect is negligible behaviorally. We discuss the welfare implications of these findings, and illustrate the quantitative differences in managerial implications between our model and the traditional one.

Key words. advertising, information, uncertainty, risk aversion, familiarity, involvement

JEL Classification: C51, D12, D80, L00, M37

1. Introduction

This study shows theoretically and empirically that exposure to advertising increases consumers' tendency to purchase the promoted product *because* the informative content of advertising resolves some of the uncertainty that the risk averse consumers face and thus reduces the risk associated with the product. We call this effect the "risk-reduction" role of advertising. The risk-reduction model implies that advertising effectiveness depends on (a) the risk preference parameter, (b) the precision of the advertising message, (c) the familiarity of the consumer with the product, (d) the diversity of products offered by multiproduct

*This study is based on the Master thesis of Dmitri Byzalov. We are grateful to Bharat Anand and Manuel Trajtenberg for helpful discussions.

firms, and (e) the consumer's sensitivity to products' attributes (and thus, her involvement level with the product).[1] These implications suggest that ads spending should be higher (a) for new and relatively unknown products, (b) for high-involvement products, (c) when ads can be quite precise, and (d) when the firm offers a diverse product-line. It also implies that ads should target consumers (a) who are more sensitive to risk, (b) who are more involved, and (c) those who are not familiar with the promoted product.

The model allows ads to affect choices also through a direct effect on the utility (i.e., the standard approach to formulate the effect of advertising). In our empirical example (where the products are television shows) the risk-reduction effect turns out to be significant and strong and the direct effect is negligible behaviorally.

The scholarly interest in economics and marketing in the avenues through which advertising affects choices has at least two major reasons. First, it is critical in constructing optimal marketing strategies. Furthermore, the issue of advertising effectiveness is gaining greater importance for several reasons, such as "... media advertising continues to draw a major proportion of the promotion budget..." and "... advertising agencies are under increasing pressure to show the specific effects of advertising on sales." (Tellis, Chandy, and Thaivanich, 2000). Second, it has welfare implications. If advertising simply has a direct effect on the utility, then ads can be claimed "... to create wants and to change and distort tastes." (Becker and Murphy, 1993).

Previous studies presented various models in order to explain how ads affect choices. The direct effect of advertising on the utility is still the dominant modeling approach (especially in empirical work).[2] Grossman and Shapiro (1984) refer to it as the "persuasive" effect of advertising, while Becker and Murphy (1993) justify it by suggesting that ads and the goods advertised can be complements. A second approach is to model advertising content as a message on products' attributes or existence. Butters (1977) and Grossman and Shapiro (1984) studied the theoretical implications of informative advertising, and Erdem and Keane (1996) and Anand and Shachar (2001) examined it empirically. A third approach demonstrates that in equilibrium ad intensity can signal product quality (Nelson, 1974; Milgrom and Roberts, 1986; Kirmani and Wright, 1989). This approach was examined empirically by Ackerberg (2003).

We build on the second approach, and show that when consumers are risk averse advertising has another avenue through which it affects choices—risk-reduction. The model is presented in Section 2. It is a choice model, in which heterogeneous consumers are facing differentiated products. They are uncertain about product attributes and have various sources of information about them. One of the sources is advertising content. Other sources are word-of-mouth, media coverage, previous experience, and the line of products offered

1 We consider a consumer who is highly (lowly) sensitive to products' attributes as high-involvement (low-involvement) consumer. The consequences on the utility of a consumer's decision are a direct function of her sensitivity to products' attributes. These consequences are significant (insignificant) for a consumer who is highly (lowly) sensitive to products' attributes. Thus, a consumer who is highly (lowly) sensitive to products' attributes is usually a high-involvement (low-involvement) consumer. This terminology is consistent with previous studies (for example, Miniard et al., 1991) and with the definition of the AMA dictionary. In any case, in this study, the degree of involvement is just an *interpretation* of consumers' sensitivity to products' attributes.

2 For example, Nevo (2000), Tellis, Chandy, and Thaivanich (2000), and Erdem and Sun (2002).

by each firm. We show that when consumers are risk averse, exposure to advertising increases their tendency to purchase the promoted product because the informative content of advertising resolves some of the uncertainty and thus reduces the risk associated with the product. In the model, we include ads not only as an element in the information set, but also allow it to have a direct effect on the utility function. This means that the model accounts for alternative effects of advertising.

We show that the risk-reduction effect of advertising has a unique structure. As mentioned above, the effectiveness of advertising (through the risk-reduction avenue) is a function of (a) the risk preference parameter, (b) the precision of the advertising message, (c) the familiarity of the consumer with the product, (d) the consumer's involvement level with the product, and (e) the diversity of products offered by multiproduct firms. In Section 2.6 we discuss the intuition behind each of these factors. For example, when a consumer faces a high-involvement product, the risk associated with the decision is high and thus the risk-reduction role of advertising is large.

Because of its unique structure and characteristics, the risk-reduction role of advertising can be distinguished empirically from the standard direct effect of ads on the utility.

The model is structurally estimated using a unique data set on television viewing choices. Accounting for the cost of leisure in consumption, television shows are clearly one of the most important consumption products. Furthermore, this data set enables us to overcome the well-known endogeneity problem of advertising exposure, as discussed in Section 4.4. An additional advantage of our empirical example is that the price of watching a television show is not product-specific. This feature is an advantage over other products, since it is occasionally difficult to track the different prices that consumers face with respect to each of the products: prices differ by firms, over time, and across consumers (through coupon schemes). Finally, almost all the commercials for television shows appear on TV. This enables us to create a comprehensive dataset of exposures to advertisements.[3] The data is presented in Section 3.

The structural estimation results (presented in Section 5) are clear: consumers are risk averse and the risk-reduction effect is significant and strong. As a result of exposure to the first ad, the average increase in the market share (via the risk-reduction effect alone) is 2.7 percentage points (which is an increase of 53.3%). On the other hand, advertising has a negligible direct effect on their utility: the average increase in the market share (via this avenue) is just 0.08 percentage points (1.5%). These results, that the risk-reduction role is strong and the direct effect is weak, are robust to various specifications of the model. Furthermore, in Section 5.4 we show that ignoring the risk-reduction role of advertising might lead to misleading implications.

As discussed above, the scholarly interest in the avenues through-which advertising affects choices has at least two major reasons: (1) it is critical in constructing optimal marketing strategies, and (2) it has welfare implications. The model sheds some light on these two issues.

3 Specifically, we obtained Nielsen individual-level panel data on television viewing choices for one week in November, 1995. We created data on show attributes, and recorded all the advertisements for these television shows—also called "previews", "promotions", or "tune-ins"—that were aired during that week. Combining our records with the Nielsen panel data and show attribute data gives us the required data to estimate the model.

The risk-reduction role of advertising has significant managerial implications. Advertising strategies are determined based on ad effectiveness. Specifically, two of the most important decisions of ad agencies depend on ad effectiveness—ads spending and audience targeting. Our model suggests that ads spending should be higher (a) for new and relatively unknown products, (b) for high-involvement products, (c) when ads can be quite precise, and (d) when the firm offers a diverse product-line. It can also assist targeting strategies. Ads should target consumers (a) who are more sensitive to risk—for example, women versus men (see Palsson, 1996; Jianakoplos and Bernasek, 1998), (b) who are more sensitive to products' attributes (and thus more involved), and (c) those who are not familiar with the promoted product.

The welfare implications of the model are quite straightforward. The concern that ads can create wants and change and distort tastes has little support *in our empirical example*, since we find that advertising has a negligible direct effect on consumers' utility. Furthermore, instead of having these negative aspects, advertising in our model informs consumers about product attributes and as a result reduces the undesirable tension that is associated with uncertainty.

1.1. Related literature

The combination of informative advertising and risk averse consumers was already presented by Erdem and Keane (1996). They observed that "the higher the advertising frequency, the higher the brand choice probability". However, unlike our model, they did not allow ads to have a direct effect on the utility. A model that assumes away the direct effect of advertising on the utility risks mistakenly detecting that ads are informative and consumers are risk averse (i.e., the risk-reduction effect). Suppose that the individuals are risk-neutral, and ads have a direct effect on the utility but they are not informative. The only way for the misspecified model to account for the positive direct effect of advertising is by attributing it to the interaction of informative ads with risk aversion in preferences, thus leading to a *spurious* finding of the risk-reduction effect. Furthermore, there are additional differences between Erdem and Keane (1996) and our study. We also (a) demonstrate how one can distinguish (empirically) between these seemingly similar effects of advertising, and (b) show that the suggested model has novel implications about advertising strategies.

The combination of informative advertising and direct effect of ads on the utility was already presented by Anand and Shachar (2001). They demonstrate empirically that advertising improves the match between consumers and products. Furthermore, they show that informative advertising might have either a positive or a negative effect on the purchasing probability.[4] They also found that advertising has a direct positive effect on the utility. We show that this finding is due to their restrictive assumption that consumers are risk-neutral. Our study, which is based on the same data, extends the empirical model of Anand and Shachar (2001) for the case of general risk preferences. Our results confirm their findings

4 Specifically, when the consumer's utility from a product is low, informative advertising decreases consumers' tendency to buy the promoted product by informing them about the product's attributes. When the consumer's utility is high advertising has a positive effect on the purchasing probability.

about the informative role of advertising. However, at the same time, we find that consumers are risk averse and the direct effect on the utility is negligible.

Our study is also related to Ackerberg (2003). Both studies distinguish between the direct and indirect effects of advertising, and both rely on the relationship between familiarity and ad effectiveness as a source of identification. However, there are several significant differences between his approach and ours. First, the indirect effect is quite different in the two studies. According to the signaling theory, people tend to buy products for which they have seen many ads *because* the perceived quality of these products is higher. According to the risk-reduction model, people tend to buy products for which they have seen many ads *because* the perceived risk of these products is lower. These differences are closely related to the fact that the risk-reduction model and the signaling theory focus on different types of products and ads. While the first applies to search goods, the second is relevant for experience goods. For search goods, ads *content* can credibly convey information on products' attributes (i.e., "this automobile has 200 horse powers"). However, for experience goods, experience is required in order to acquire information about the utility, and ads' content cannot credibly convey information about product quality. Thus, according to Nelson (1974), ad *intensity* (not ad content) serves as a signal about the quality of the product. Furthermore, in Section 5.5 we show that the relationship between advertising intensity and product quality in our data is not consistent with the signaling mechanism of informative advertising. This allows us to rule out the signaling hypothesis for our empirical example. Second, unlike in Ackerberg (2003), variation in familiarity is not the only source of identification in our set-up. Our additional sources of identification are variation in involvement across consumers, and variation in the line of products across firms. Thus, our approach allows to distinguish between the direct effect and the informative effect even if there is no variation in familiarity. Furthermore, we use a somewhat different source of heterogeneity in familiarity for identification. As demonstrated by Shachar and Anand (1998), the informative role of advertising can be identified using variation in familiarity across products. Ackerberg (2003) used variation in familiarity across consumers, rather than across products. We combine the two approaches and use variation across both products and consumers.

This paper is structured as follows. In Section 2, we present the model and discuss the model implications that allow for empirical identification. In Section 3, we present the dataset. In Section 4, we specify the details of the empirical model, and outline the identification scheme. Section 5 summarizes the findings. Section 6 concludes and presents managerial implications.

2. The model

We start by describing the setting of the model, the utility function, and the information set of the individual. Then, we present several implications of the model that demonstrate the risk-reduction role of advertising. These unique implications enable us to empirically distinguish between the direct effect of ads on the utility and the risk-reduction effect.

The key ingredients of the model and its implications are not industry-specific. However, applying the model to the television example requires accounting for the specificity of viewing behavior. Thus, in order to make the presentation clear, we describe the model in the context of the empirical example.

The example is on television viewing choices in the US for the week that starts on November 6, 1995. At that time there were 4 major television networks: ABC, CBS, NBC and FOX. We focus on prime-time shows (Monday through Friday, 8:00–10:00 PM). Our data, presented in Section 3, consists of (a) individual level panel of choices, (b) individuals' characteristics, (c) individuals' exposure to advertising, and (d) products' attributes. We observe choices for every 15 minutes. Thus, for each individual we observe 40 viewing choices. Notice that in this example, the products are television shows, an ad for a product is usually referred to as a 'promo', and a period is termed a 'time slot'.

2.1. *The setup*

The timing and information structure of the model are tailored for our empirical example of television viewing. In particular, there are J firms (the 4 leading television networks in the empirical example) in the market, and T periods (40 quarter hour time slots during prime time). In period t each firm offers one product (airs a television show). Each product is offered for several consecutive periods (2 to 8 time slots in the empirical example), after that the firm switches to its next product.

There are I individuals who are indexed by i. They face $J+1$ mutually exclusive and exhaustive alternatives, corresponding to: (0) Outside, (1) ABC, (2) CBS, (3) NBC, and (4) FOX. In each period t, individual i makes a choice, $C_{i,t}$, from among these $J+1$ options indexed by j. Thus, $C_{i,t}=j$ when individual i chooses alternative j at time t.

2.2. *The utility*

2.2.1. The utility from watching network television. The utility of individual i from the product offered by firm j in period t is:

$$U_{i,j,t} = V(x_{j,t}\beta_i + \eta_{j,t}) + \rho_i\left(N^a_{i,j,t}\right) + \delta_{i,j,t} I\{C_{i,t-1} = j\} + \alpha_{i,j} + \nu_{i,j,t} + \varepsilon_{i,j,t} \quad (1)$$

The element $x_{j,t}\beta_i + \eta_{j,t}$ represents the match between the observed product attributes, $x_{j,t}$, and the individual's preferences, β_i. The variable $x_{j,t}$ is a K-dimensional row-vector, and the parameter β_i is a K-dimensional vector. The parameter vector β_i is a function of observed and unobserved individual characteristics. The parameter $\eta_{j,t}$ can be thought of as the mean (across individuals) of the unobserved interactions between products' attributes and consumers' tastes.[5] Henceforth, we will denote the element $x_{j,t}\beta_i + \eta_{j,t}$ by $z_{i,j,t}$, and refer to it as "attribute utility" (since it captures the match between product attributes and individual preferences). The curvature of $V(\cdot)$ captures the individual risk preferences.

5 (1) In the industrial organization literature the element $x_{j,t}\beta_i$ is called "the horizontal dimension of utility", and $\eta_{j,t}$ "the vertical dimension".
(2) The $\eta_{j,t}$ parameter is fixed for the duration of each show. Consequently, a half-hour show and a one-hour movie each have one η parameter.

The second term $\rho_i(N^a_{i,j,t})$ captures the direct effect of advertising on the utility, where the variable $N^a_{i,j,t}$ is the number of advertisements that individual i was exposed to with respect to the show aired by network j at time t. The direct effect was termed 'persuasive' by Grossman and Shapiro (1984). Although there might be other interpretations for this effect, we would refer to it as 'persuasive' for simplicity of the presentation. The persuasive effect of advertising implies that exposing the individual to $N^a_{i,j,t}$ ads for product j, t directly increases her utility from that product (i.e., $\frac{\partial \rho_i}{\partial N^a_{i,j,t}} > 0$).

The third term $\delta_{i,j,t} I\{C_{i,t-1} = j\}$ captures behavioral state-dependence, i.e. the utility from the alternatives available in the current period t may depend on the individual's choice in the previous period. The indicator function $I\{\cdot\}$ is equal to one if the individual purchased the product offered by firm j in the previous period, and to zero otherwise. The parameter $\delta_{i,j,t}$ is a function of observable and unobservable individual characteristics, product attributes, and time. There are various sources for state dependence: habit persistence, switching costs, asymmetric information and search costs (Moshkin and Shachar, 2002), and learning (from past experiences) that reduces uncertainty (Erdem, 1998). Previous studies of television viewing choices find strong evidence of state dependence even when unobserved heterogeneity is accounted for.[6]

The last three terms in the utility ($\alpha_{i,j}$, $\nu_{i,j,t}$ and $\varepsilon_{i,j,t}$) are unobserved by the researcher. The parameter $\alpha_{i,j}$ represents an unobserved network-individual match. It is common to all the products offered by firm j. The random variable $\nu_{i,j,t}$ represents the show-individual random effect, common to all the time slots of a show, but independent across shows and individuals. Since $\alpha_{i,j}$ and $\nu_{i,j,t}$ are unobserved by the researcher, ignoring them can bias the estimates of the state dependence parameters $\delta_{i,j,t}$ (Heckman, 1981). The random variable $\varepsilon_{i,j,t}$ captures transitory effects, assumed to be independent across time periods, products and individuals. Notice that $\nu_{i,j,t}$ does not change during the duration of a show, while $\varepsilon_{i,j,t}$ does.

2.2.2. The utility from the outside alternative. The outside utility is a function of the individual's characteristics and state dependence. Specifically, it is:

$$U_{i,0,t} = \gamma_i + \delta_{i,0,t} I\{C_{i,t-1} = 0\} + \varepsilon_{i,0,t} \tag{2}$$

where γ_i is allowed to depend on the observable and unobservable individual characteristics. The term $\delta_{i,0,t} I\{C_{i,t-1} = 0\}$ captures behavioral state-dependence. The transitory random term $\varepsilon_{i,0,t}$ is i.i.d. across time periods and individuals.

6 In the television industry this well known phenomenon is called the "lead-in effect." Darmon (1976) introduces the concept of channel loyalty and Horen (1980) estimates a lead-in effect, both using aggregate ratings models. Rust and Alpert (1984) use individual-level data to estimate an audience flow model, in which viewers are described as being in one of five states according to: whether the television was previously on or off; if it was on, whether it was tuned to the same channel as the current viewing option; and whether this option is the start or continuation of a show. Shachar and Emerson (2000) allow state dependence to vary across shows and across demographically defined viewer segments. Goettler and Shachar (2001) demonstrate that the cost of switching remains when unobserved heterogeneity is accounted for.

2.3. Information set

We assume that the consumers do not have perfect information on the product attributes $(x_{j,t}, \eta_{j,t})$, and therefore they face uncertainty about the actual "attribute utility" $z_{i,j,t} = x_{j,t}\beta_i + \eta_{j,t}$. A priori, the assumption of imperfect information seems appropriate for most consumer goods, due to substantial product differentiation, large number (and rapid introduction) of brands and products, as well as cognitive limitations or costs of verifying or remembering product characteristics. For illustration, although the data in our empirical example span only 10 hours of prime-time television (5 days and 2 hours a day), the 4 leading television networks alone offered (in this time frame) 53 different products. Thus, imperfect information and reliance on different sources of information could be essential in our empirical example. In the empirical study we test this hypothesis (i.e., the model nests perfect information as a special testable case).

A consumer's information set includes (a) a prior distribution of products' attributes, and (b) product-specific signals such as advertising and word-of-mouth.

2.3.1. The prior distribution. In many markets, multi-product firms tend to specialize in products with a distinctive set of characteristics. Consumers are usually aware of the specialization of each of the firms. The "profile" of the multi-product firms can convey useful information to a consumer facing uncertainty about product characteristics. For example, knowing that a particular car is a Toyota can be helpful in inferring reliability. Indeed, Anand and Shachar (2004) demonstrate, using the same data set as we use, that the profiles of the multi-product firms are an important element in the information set of consumers. We follow their formulation and assume that both $\eta_{j,t}$ and $x_{j,t}$ follow a normal distribution, and thus the prior distribution of individual i on $z_{i,j,t}$ is:

$$z_{i,j,t} \sim N\left(\mu_{i,j}, \left[\varsigma_{i,j}^{p}\right]^{-1}\right) \tag{3}$$

where $\varsigma_{i,j}^{p}$ is the precision (reciprocal of the variance) of the prior distribution, and by definition,

$$\mu_{i,j} = E_t[\eta_{j,t}] + E_t[x_{j,t}]\beta_i \tag{4}$$

where $E_t[\cdot]$ is the expected value across time slots. Hereafter, we refer to $\mu_{i,j}$ as brand image. We assume that consumers are aware of the specialization of each of the firms (i.e., consumers know $E_t[\eta_{j,t}]$, $E_t[x_{j,t}]$, and the variance-covariance matrix of these attributes). In other words, while the individual is uncertain about product attributes, she knows the profile of each firm.[7]

2.3.2. The product-specific signals. Consumers receive information on specific products from various sources: word-of-mouth, media coverage, previous experience, and advertising. We assume that this information is noisy and formulate it as product-specific

7 Thus, in the empirical example, $\mu_{i,j} = \frac{1}{T}\sum_t z_{i,j,t}$, and $\varsigma_{i,j}^{p} = [\frac{1}{T-1}\sum_t (z_{i,j,t} - \mu_{i,j})^2]^{-1}$.

signals. Specifically, we formulate each of the k unbiased noisy signals $S_{i,j,t,k}$ as:

$$S_{i,j,t,k} = z_{i,j,t} + \omega_{i,j,t,k} \quad \text{where} \quad \omega_{i,j,t,k} \sim N\big(0, \varsigma_{i,j,t,k}^{-1}\big) \tag{5}$$

and where $\varsigma_{i,j,t,k}$ is the precision of each signal.

Miscellaneous signals. Even an individual who was not exposed to any ads has some information about products. Her knowledge is based on various sources such as word-of-mouth or media coverage. In the case of television shows, the viewer can also obtain information from the schedule published in the newspapers or from watching previous episodes of the same show.[8]

We lump all these sources of information under the title "miscellaneous signals". Since we do not observe the number of miscellaneous signals received by each individual, we normalize this number to be one. However, we allow the precisions of these signals (denoted by $\varsigma_{i,j}^{m}$) to vary across individuals and firms.[9] This means that some people could be better informed than others, and the products of some brands could be more familiar than others.

The extreme case of $\frac{1}{\varsigma_{i,j}^{m}} = 0$ corresponds to perfect information on product characteristics.[10]

Advertising. We formulate ads as noisy signals with precision ς^{a}. The noisiness of advertising is well-documented.[11] We assume that the ad signals are independent for two reasons: (1) firms occasionally use different advertisements for the same product; (2) different exposures to the same advertisement can lead to different impressions. The independence assumption does not affect our qualitative results.[12]

The effect of advertisements through the information set is determined by ς^{a}. If $\varsigma^{a} = 0$, then advertisements are too noisy to convey any information about product attributes, i.e. advertising exposures have no effect on the information set of the consumer. On the other hand, when $\varsigma^{a} > 0$, each advertising exposure affects the information set. Thus, ς^{a} is one of the parameters of interest in the empirical study.

Experience. A feature unique to our data set is that each product is being consumed over several consecutive periods. Since the product characteristics $(x_{j,t}, \eta_{j,t})$ do not vary

8 Our data span one week of television viewing, so the exposures to previous episodes of the show are unobserved.

9 This is equivalent to fixing the variance of all the signals and estimating the number of signals.

10 The miscellaneous signals are independent across products and individuals, however the same realization of the miscellaneous signal applies to all the time periods in which the product is being offered.

11 See, for example, Jacoby and Hoyer (1982). Using a survey of 2,700 consumers about the content of 60 thirty-second televised communications (including advertisements), they find that 29% of these were miscomprehended by consumers. They find similar results in their 1989 study, which uses a survey of 1,250 consumers who were exposed to print ads.

12 (1) The unbiasedness assumption rests on truth-in-advertising regulations. Furthermore, if a firm has an incentive to bias the content of its advertisements, a rational consumer would account for it, and is likely to neutralize the bias. We do not model this game in order to keep the model focused on its key elements.
(2) Note that although the advertising signals are independent across individuals and products, they are the same for any period of each product. In the model, each product is offered for several consecutive periods, and the product characteristics are assumed to be constant throughout these periods.

across these periods, the individual's perception of the product attributes becomes more accurate while consuming it. Formally, each consumption period provides the individual with an additional "experience signal" with a precision ς^e. Let $N^e_{i,j,t}$ represent the number of periods that individual i watched the current show on network j prior to period t.[13]

2.4. *The posterior "attribute utility"*

The posterior "attribute utility" depends on the realizations of all the signals, $S_{i,j,t,k}$. It is equal to (see DeGroot, 1989):

$$\mu^p_{i,j,t} = \mu_{i,j} + \theta_{i,j,t}(z_{i,j,t} - \mu_{i,j}) + \vartheta_{i,j,t}\omega_{i,j,t} \tag{6}$$

where

$$\theta_{i,j,t} = \left[\varsigma^m_{i,j} + \varsigma^a N^a_{i,j,t} + \varsigma^e N^e_{i,j,t}\right]\left[\varsigma^p_{i,j} + \varsigma^m_{i,j} + \varsigma^a N^a_{i,j,t} + \varsigma^e N^e_{i,j,t}\right]^{-1}, \tag{7}$$

$$\vartheta_{i,j,t} = \left[\sqrt{\varsigma^m_{i,j}} + \sqrt{\varsigma^a N^a_{i,j,t}} + \sqrt{\varsigma^e N^e_{i,j,t}}\right]\left[\varsigma^p_{i,j} + \varsigma^m_{i,j} + \varsigma^a N^a_{i,j,t} + \varsigma^e N^e_{i,j,t}\right]^{-1}, \tag{8}$$

and where $\omega_{i,j,t}$ is a standard normal deviate, independent across products and individuals, but serially correlated within the time periods of each product.[14]

The intuition of this equation is the following. If the individual does not receive any product specific signal then $\theta_{i,j,t} = \vartheta_{i,j,t} = 0$ and the posterior "attribute utility" is simply the mean of the prior, $\mu_{i,j}$. When she gets such signals, she updates her prior based on the signals. The researcher does not observe the realizations of the signals, but unlike the individual the researcher observes the product attributes and thus $z_{i,j,t}$. The update of the prior (on average) is toward the actual "attribute utility", $z_{i,j,t}$ which is the expected value of the signals. The magnitude of the update is an increasing function of both the precision and the number of the product specific signals. The last element in equation (6) accounts for the noisiness of the signals.

Both $(1 - \theta_{i,j,t})$ and $\vartheta_{i,j,t}$ can be viewed as measures of the consumers' knowledge on the specific product. For example, in the extreme case when the miscellaneous signals give precise information ($\frac{1}{\varsigma^m_{i,j}} = 0$), we get that $(1 - \theta_{i,j,t}) = \vartheta_{i,j,t} = 0$, and the posterior mean, $\mu^p_{i,j,t}$, is equal to the actual "attribute utility", $z_{i,j,t}$. The informative role of advertising expresses itself through equations (6)–(8). As $N^a_{i,j,t}$ increases, $(1 - \theta_{i,j,t})$ and $\vartheta_{i,j,t}$ decrease, and as a result the posterior "attribute utility" is (on average) closer to $z_{i,j,t}$, while the variance of the error term in the posterior declines toward zero. The magnitude of these changes depends on the precision of the advertising signals, ς^a. While this effect was already pointed out by Anand and Shachar (2001), this model presents an additional role for advertising, presented in the following sub-section.

13 No experience signals are observed for the alternatives not chosen in the current period. The signals are independent across individuals, products and time periods.

14 This correlation arises due to our specification that the same realization of advertising and miscellaneous signals applies to all the time periods of the product, and due to the accumulation of experience signals.

2.5. The posterior variance of the "attribute utility"

The expected utility of a risk averse individual is not only a function of the posterior mean of the "attribute utility", $\mu^p_{i,j,t}$, but also a function of its posterior variance, denoted by $\sigma^2_{i,j,t}$.[15] Let $\tilde{V}(\mu^p_{i,j,t}, \sigma^2_{i,j,t})$ represent the expected value of $V(z_{i,j,t})$.

The posterior variance of the "attribute utility" is:

$$\sigma^2_{i,j,t} = \left[\varsigma^p_{i,j} + \varsigma^m_{i,j} + \varsigma^a N^a_{i,j,t} + \varsigma^e N^e_{i,j,t}\right]^{-1} \tag{9}$$

It is a decreasing function of $N^a_{i,j,t}$. If consumers are risk-neutral, then their expected utility does not depend on $\sigma^2_{i,j,t}$. However, if they are risk-averse, exposure to advertising has another effect on their expected utility: as $N^a_{i,j,t}$ increases, $\sigma^2_{i,j,t}$ decreases and the expected utility increases. This is the risk-reduction role of advertising. The following sub-sections describe this effect and its implications.

2.6. The risk reduction role of advertising (theory)

The following derivative summarizes the three effects that exposure to advertising has on the expected utility (and hence on choices).

$$\begin{aligned}\frac{\partial E(U_{i,j,t}|S_{i,j,t})}{\partial N^a_{i,j,t}} &= \frac{\partial \tilde{V}\left(\mu^p_{i,j,t}, \sigma^2_{i,j,t}\right)}{\partial \mu^p_{i,j,t}} \frac{\partial \mu^p_{i,j,t}}{\partial N^a_{i,j,t}} \\ &\quad + \frac{\partial \tilde{V}\left(\mu^p_{i,j,t}, \sigma^2_{i,j,t}\right)}{\partial \sigma^2_{i,j,t}} \frac{\partial \sigma^2_{i,j,t}}{\partial N^a_{i,j,t}} + \frac{\partial \rho_i\left(N^a_{i,j,t}\right)}{\partial N^a_{i,j,t}}\end{aligned} \tag{10}$$

The first effect, $\frac{\partial \tilde{V}(\mu^p_{i,j,t},\sigma^2_{i,j,t})}{\partial \mu^p_{i,j,t}} \frac{\partial \mu^p_{i,j,t}}{\partial N^a_{i,j,t}}$, was already introduced by Anand and Shachar (2001). They have demonstrated that this effect can be either positive or negative. More specifically, the sign of the first element depends on the sign of $(z_{i,j,t} - \mu_{i,j})$.

The last effect, $\frac{\partial \rho_i(N^a_{i,j,t})}{\partial N^a_{i,j,t}}$, captures the traditional role of advertising. That is, exposure to advertising increases the tendency to purchase the product by directly increasing the utility from the product. The contribution to the utility might have diminishing returns (i.e., the "wear-out" effect).[16]

The second effect, $\frac{\partial \tilde{V}(\mu^p_{i,j,t},\sigma^2_{i,j,t})}{\partial \sigma^2_{i,j,t}} \frac{\partial \sigma^2_{i,j,t}}{\partial N^a_{i,j,t}}$, represents the risk-reduction role of advertising. Notice that $\frac{\partial \sigma^2_{i,j,t}}{\partial N^a_{i,j,t}} = \frac{-\varsigma^a}{[\varsigma^p_{i,j}+\varsigma^m_{i,j}+\varsigma^a N^a_{i,j,t}+\varsigma^e N^e_{i,j,t}]^2}$. This effect has a very unique structure. First, it, obviously, depends on the risk-parameter of the individual (which determines $\frac{\partial \tilde{V}(\mu^p_{i,j,t},\sigma^2_{i,j,t})}{\partial \sigma^2_{i,j,t}}$). In the extreme case of risk-neutral consumers (i.e., $\frac{\partial \tilde{V}(\mu^p_{i,j,t},\sigma^2_{i,j,t})}{\partial \sigma^2_{i,j,t}} = 0$) this effect is irrelevant.

15 Notice that since the "attribute utility" is normally distributed, the expected utility depends only on its first two moments.

16 In other words, it is usually assumed (and found in empirical studies) that the first derivative ($\frac{\partial \rho_i(N^a_{i,j,t})}{\partial N^a_{i,j,t}}$) is positive and the second ($\frac{\partial^2 \rho_i(N^a_{i,j,t})}{(\partial N^a_{i,j,t})^2}$) is negative.

On the other hand, the risk-reduction will be stronger when the consumers are more risk-averse.

Second, the risk-reduction depends on the precision of the advertising message. When $[\varsigma^{p}_{i,j} + \varsigma^{m}_{i,j} + \varsigma^{e} N^{e}_{i,j,t}] > \varsigma^{a} N^{a}_{i,j,t}$ (as is in our estimation results) the higher ς^{a}, the stronger the effect. The rationale of this result goes as follows: as ads become more precise they resolve the uncertainty (and the associated risk) that the consumer is facing more effectively.

Third, it depends on the familiarity of the consumer with the specific product. As the consumer knows more about the product a-priori (i.e., $\varsigma^{m}_{i,j}$ is higher), the risk-reduction effect is weaker. The intuition of this is simple: advertising provides information and thus reduces uncertainty. If the consumer is quite familiar with the product, the uncertainty (and the associated risk) that she is facing is small to begin with.

Fourth, the risk-reduction is a function of the consumer's sensitivity to products' attributes (the scale of β_i) which we interpret as the level of the consumer's involvement with the product. It is easy to show that if β_i is larger (in absolute terms) $\varsigma^{p}_{i,j}$ is smaller and as a result, $\frac{\partial \sigma^2_{i,j,t}}{\partial N^a_{i,j,t}}$ is larger. The intuition of this result is the following: consumers who are highly involved with a product are facing a higher risk than those whose involvement level is low. Thus, for such consumers, the risk-reduction role of advertising is more salient.

Fifth, the risk-reduction effect depends on the diversity of products offered by a multiproduct firm. The larger the variety of products (i.e., $\mathrm{Var}(x_{j,t})$ is large for firm j), the stronger the risk-reduction effect. It is easy to show that if $\mathrm{Var}(x_{j,t})$ is larger, $\varsigma^{p}_{i,j}$ is smaller and as a result, $\frac{\partial \sigma^2_{i,j,t}}{\partial N^a_{i,j,t}}$ is larger. The rationale behind this result is the following: the consumer uses the multiproduct firm's profile to resolve some of the uncertainty that she is facing. The larger the variety of products, the less precise is this source of information, and, as a result, the higher the uncertainty (and the associated risk) that the consumer is facing.

To summarize: the risk-reduction role of advertising is a function of: (a) the risk preferences parameter, (b) the precision of the advertising message, (c) the familiarity of the consumer with the product, (d) the consumer's involvement level with the product, and (e) the diversity of products offered by the multiproduct firms.

These results have various managerial implications that we discuss in Section 6. For example, the results imply that advertising effectiveness is large for new or unknown products, and for high-involvement products.

Furthermore, the unique structure of the risk-reduction effect clarifies that it is possible to empirically distinguish between it and the traditional direct effect on the utility. Specifically, unlike the risk-reduction effect, the direct effect on the utility does not depend on factors such as the familiarity of the consumer with the product, her sensitivity to product attributes (our measure of involvement), and the diversity of products offered by the firm. This allows us to distinguish between the risk-reduction effect and the direct effect, for very general specifications of the direct effect.[17] The combination of rich data and novel implications augments the identification scheme compared to previous studies. Previous

17 For example, in the estimation we allow the direct effect to depend on the individuals' observable and unobservable characteristics, and allow for non-parametric specifications.

efforts to distinguish between the direct and the indirect effect of advertising based their identification scheme on the differences in the familiarity across products (Shachar and Anand, 1998) or consumers (Ackerberg, 2003). We combine the two approaches and use variation across both products and consumers. Moreover, the identification of this model is richer. It is not based only on the variation in the familiarity, but also on the involvement level of the consumer with the product, the diversity of products offered by the firm, etc.

3. The 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-level exposures to advertisements (which promote 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 empirical analysis of advertising. We focus on viewing choices for network television during prime time, 8:00 to 10:00 PM–Monday through Friday starting on November 6, 1995. There are 53 shows during this time-frame.

The advantages of this data set are: (1) previous studies demonstrate that TV viewers are uncertain about product attributes, (2) Anand and Shachar (1998 and 2001) show that ads for TV shows are informative, (3) most ads for these products are observed, since most ads for television shows appear on TV, (4) the monetary cost of viewing a show is zero, and the non-monetary cost is the same (for each individual) across shows in any period—thus, we do not face the complications due to manufacturer coupons or special offers, and (5) it is relatively easy to solve the endogeneity problem of the advertising exposure (as discussed in Section 4.4).

3.1. The data sets

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

3.1.1. Product (show) characteristics. We coded the show attributes for the 53 shows in the relevant week based on prior knowledge, publications about the shows, and viewing each one of them. Following previous studies, we categorize shows based on their genre and their cast demographics. Rust and Alpert (1984) present five show categories—for example, comedies and action dramas—and show that viewers differ in their preferences over these categories. We use the following categories: *situation comedies*, also called “sitcoms” (31 shows fall into this category), *action dramas* (11 shows), and *romantic dramas* (7 shows). The base group includes news magazines and sports events (4 shows), which were found by previous studies to be similar.[18]

18 See Goettler and Shachar (2001).

Shows were also characterized by their cast demographics. Shachar and Emerson (2000) demonstrate that the demographic match between an individual and a show's cast plays an important role in determining viewing choices. For example, younger viewers tend to watch shows with a young cast, while older viewers prefer an older cast. We use the following categories: *Generation-X*, if the main characters in a show are older than 18 and younger than 34 (21 shows fall into this category); *Baby Boomer*, if the main 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).

3.1.2. 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. Nielsen installs a People Meter (NPM) for *each* television set in the household. The NPM records the channel being watched on each television set. A special remote-control records the individuals watching each TV. Thus, the viewing choices are *individual-specific*. While criticized occasionally by the networks, Nielsen data still provide the standard measure of ratings for both network executives and advertising agencies.

Although the NPM is calibrated for measurements each minute, the data available to us provide quarter-hour viewing decisions, measured as the channel being watched at the midpoint of each quarter-hour block. Thus, we observe viewers' choices in 40 time slots. Our data consists of viewing choices for the four major networks, ABC, CBS, NBC, and FOX. The length of the shows varies from 30 minutes to 2 hours.

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 presents their summary statistics.

3.1.3. Data on exposures to advertising. We taped all the shows for the four networks during the relevant week, and coded the appearance of each advertisement for the television shows. For example, on Tuesday at 8:00 PM, there was an advertisement for the NBC sitcom *Seinfeld* (this show aired on Thursday at 9:00 PM). This information was matched with the Nielsen viewing data to determine an individual's exposure to advertisements. For example, an individual who watched NBC on Tuesday at 8:00 PM was exposed to the advertisement

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.

Table 1. Individual observable characteristics: definitions and summary statistics.

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

mentioned above. Summing over all time slots, we get the number of exposures of individual i with respect to each show in the week. In 1995, these advertisements, which are also referred to as "promos", usually included the broadcast time of the show, and clips from the actual episode.

Since our Nielsen viewing data starts on Monday we cannot determine the exposure to advertisements that were aired before that day. This means that our data miss some ad exposures for the shows in the relevant week. 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, we allow the advertising parameters to differ across these two parts of the week.

For the Wednesday through Friday shows, the mean number of advertisements aired per show is 4.14, and the median is 4. On average, an individual who watched TV at least 30 minutes during Monday and Tuesday is exposed to 0.49 advertisements for each show on Wednesday through Friday.

4. Estimation and identification issues

This section consists of four subsections. The first presents the specific functional forms of the utility, and the density functions of the unobserved variables. The second constructs the likelihood function. The identification of the model's parameters is discussed in the third subsection, and the final subsection analyzes the endogeneity problem.

4.1. The functional forms

4.1.1. Utility. *Attribute utility:* Following previous studies we model the "attribute utility" from television show as:

$$\begin{aligned} z_{i,j,t} = {} & \beta_{Gender} I\{\text{the gender of } i \text{ and show } j,t \text{ is the same}\} \\ & + \beta_{Age0} I\{\text{the age group of } i \text{ and show } j,t \text{ is the same}\} \\ & + \beta_{Age1} I\{\text{the distance between the age group of } i \text{ and show } j,t \text{ is one}\} \\ & + \beta_{Age2} I\{\text{the distance between the age group of } i \text{ and show } j,t \text{ is two}\} \\ & + \beta_{Family} I\{i \text{ lives with her family and show } j,t \text{ is about family matters}\} \\ & + \beta_{Race} Income_i I\{\text{one of the main characters in show } j,t \text{ is} \\ & \qquad\qquad \text{African American}\} \\ & + \left(\beta_{Sitcom} \cdot y_i^{\beta} + \upsilon_i^{Sitcom}\right) Sitcom_{j,t} \\ & + \left(\beta_{AD} \cdot y_i^{\beta} + \upsilon_i^{AD}\right) ActionDrama_{j,t} \\ & + \left(\beta_{RD} \cdot y_i^{\beta} + \upsilon_i^{RD}\right) RomanticDrama_{j,t} + \eta_{j,t} \end{aligned} \tag{11}$$

The first six terms in $z_{i,j,t}$ capture the match between the demographics of the viewer and the main characters in the cast, with $Income_i$ used as a proxy for the individual's race. The binary variables $Sitcom_{j,t}$, $ActionDrama_{j,t}$, $RomanticDrama_{j,t}$ describe the genre of show j, t.[20] The preferences for the show genre are allowed to depend on the individual observable (y_i^{β}) and unobservable (υ_i^{Sitcom}, υ_i^{AD}, υ_i^{RD}) characteristics. The vector y_i^{β} consists of the variables $Teens_i$, $GenerationX_i$, $BabyBoomer_i$, Old_i, $Female_i$, $Income_i$, $Education_i$, $Family_i$ and $Urban_i$.

The direct (persuasive) effect of advertising: The ρ function has the form:

$$\rho_i\left(N_{i,j,t}^a\right) = \upsilon_i^{\rho}\left[\rho_{1,t} N_{i,j,t}^a + \rho_{2,t}\left(N_{i,j,t}^a\right)^2\right] \tag{12}$$

where $\rho_{1,t} = \rho_{1,MT} MT_t + \rho_{1,WF} WF_t$, and MT_t and WF_t are binary variables for Monday-Tuesday and Wednesday-Friday respectively.[21] The parameter $\rho_{2,t}$ is defined accordingly. The response to ad exposure is allowed to vary across consumers via the unobserved parameter υ_i^{ρ}. The "wear-out" effect of advertising implies that $\rho_{2,t} < 0$.

The state dependence parameter: Next, 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:

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

20 Following the findings of Goettler and Shachar (2001), we merge News and Sports shows into a single baseline category.

21 Recall that we do not observe if the individuals were exposed to ads prior to Monday. Thus, we have missing exposure data. This problem is likely to be more severe for the shows aired in the first days of our dataset.

where

$$\delta_{i,j,t} = \begin{bmatrix} y_i^{\delta}\delta^{y} + \upsilon_i^{\delta} \\ + \delta_{First15} I\{\text{The show on } j \text{ started within the past 15 minutes}\} \\ + \delta_{Last15} I\{\text{The show on } j \text{is at least one hour long and will end} \\ \text{within 15 minutes}\} \\ + x_{j,t}\delta^{x} I\{\text{The show on } j \text{ started at least 15 minutes ago}\} \end{bmatrix}$$

where the observed variables included in y_i^{δ} are $Teens_i$, $GenerationX_i$, $BabyBoomer_i$, $Older_i$, and $Female_i$. We also allow the state dependence to differ across individuals for unobserved reasons through υ_i^{δ}, and over time through $\delta_{First15}$, δ_{Last15}. Specifically, we expect δ to be smaller in the first 15 minutes of a show, when viewers have not had enough time to get hooked by the show ($\delta_{First15} < 0$). For the same reason, the state dependence should be particularly high during the last 15 minutes of a show ($\delta_{Last15} > 0$). Furthermore, we allow the state dependence during the show to depend on the show characteristics $x_{j,t}$ (note that δ^{x} does not apply between shows). 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 at the beginning of a new show, $\delta_{InProgress}$ is expected to be negative.

Note that our data span 5 days, with 8 observations per day. The state dependence term applies to consecutive observations within each day, but is set to zero on transitions between days.

Outside utility: The parameters of the outside utility are defined as:

$$\begin{aligned} \gamma_i &= \gamma y_i^{\gamma}, \text{ and} \\ \delta_{i,0,t} &= \delta_0 + \upsilon_i^{\delta} + \delta^{Y} y_i^{\delta} + \delta_{Hour} Hour_t \end{aligned} \tag{13}$$

where the vector y_i^{γ} includes all the demographic variables from y_i^{β}, as well as binary variables for the individual cable subscription status $Basic_i$, $Premium_i$, and the binary variable $Hour_t$ equals 1 at 8:00 and 9:00 PM. Notice that the outside alternative lumps together the TV off choice with the decision to watch a non-network show. Thus, we expect the cable subscription status to affect the utility from the outside alternative. Furthermore, we include the variable $Hour_t$ in the state dependence parameter, since most non-network shows end "on the hour".

Risk aversion: We formulate the utility from the product attributes so that:

$$\tilde{V}\left(\mu_{i,j,t}^{p}, \sigma_{i,j,t}^{2}\right) = \mu_{i,j,t}^{p} + \phi\sigma_{i,j,t} \tag{14}$$

In this semi-structural formulation, the case $\phi = 0$ corresponds to risk neutrality, and $\phi < 0$ corresponds to risk aversion.

The most natural alternative specification would be to capture risk preferences via the curvature of $V(\cdot)$. However, this approach is more sensitive empirically. In particular, if the functional form of $z_{i,j,t}$ or the distribution of $\nu_{i,j,t}$, $\varepsilon_{i,j,t}$ is misspecified, then the estimated shape of $V(\cdot)$ could be driven by this misspecification, rather than by the effect of posterior variance on the expected utility (which is our desired source of identification for the parameters of risk preferences). Therefore we choose to use the more robust semi-structural specification.

Furthermore, using Taylor expansion, $\tilde{V}(\mu^p_{i,j,t}, \sigma^2_{i,j,t})$ can be re-written as:

$$\begin{aligned}\Delta\tilde{V}(\mu,\sigma) = {} & \frac{\partial\tilde{V}}{\partial\mu}\Delta\mu + \frac{\partial\tilde{V}}{\partial\sigma}\Delta\sigma + \frac{1}{2}\frac{\partial^2\tilde{V}}{\partial\mu^2}(\Delta\mu)^2 \\ & + \frac{1}{2}\frac{\partial^2\tilde{V}}{(\partial\sigma)^2}(\Delta\sigma)^2 + \frac{1}{2}\frac{\partial^2\tilde{V}}{\partial\mu\partial\sigma}\Delta\mu\Delta\sigma + \cdots \end{aligned} \tag{15}$$

Note that under risk-neutrality $\frac{\partial\tilde{V}}{\partial\sigma} = 0$, and under risk-aversion $\frac{\partial\tilde{V}}{\partial\sigma} < 0$. In preliminary stages of estimation, we experimented with different orders of Taylor approximation, and found that a simple first-order approximation works best empirically.

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

Let υ_i represent all the unobserved individual specific parameters. Specifically, $\upsilon_i = \{\alpha_{i,j},\ \upsilon_i^{Sitcom},\ \upsilon_i^{AD},\ \upsilon_i^{RD},\ \upsilon_i^{\rho},\ \upsilon_i^{\delta},\ \varsigma^m_{i,j}\}$. Without accounting for υ_i, our estimates of the parameters of the state dependence, the advertising effectiveness and the information set might be inconsistent (Heckman, 1981; Shachar and Anand, 1998).

The density function of υ_i is assumed to be discrete. Specifically, $\upsilon_i = \upsilon_k$ with probability p_k, where $p_k = \frac{\exp(\lambda_k)}{\sum_{k=1}^{K}\exp(\lambda_k)}$ 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.

Finally, the density function of the individual-show match unobserved variable, $\nu_{i,j,t}$, is assumed to be normal – $\nu_{i,j,t} \sim N(0, \sigma^2_\nu)$.

4.2. *The likelihood*

The choice probability in period t is:

$$\Pr\left(C_{i,t} = j \mid C_{i,t-1}, x_t, y_i, N^a_{i,t}, N^e_{i,t}, \omega_{i,t}, \nu_{i,t}, \upsilon_k, \theta\right) = \frac{\exp(\bar{U}_{i,j,t})}{\sum_{m=0\ldots J}\exp(\bar{U}_{i,m,t})} \tag{16}$$

where $\omega_{i,t}$, $\nu_{i,t}$, x_t, $N^a_{i,t}$ and $N^e_{i,t}$ are the J-element vectors whose j'th component is $\omega_{i,j,t}$, $\nu_{i,j,t}$, $x_{j,t}$, $N^a_{i,j,t}$ and $N^e_{i,j,t}$ respectively; θ a vector of all the parameters that are common for all individuals; $\bar{U}_{i,j,t} = \mu^p_{i,j,t} + \phi\sigma_{i,j,t} + \rho_i(N^a_{i,j,t}) + \delta_{i,j,t}I\{C_{i,t-1} = j\} + \alpha_{i,j} + \nu_{i,j,t}$ for $j = 1, .., J$, and $\bar{U}_{i,0,t} = \gamma_i + \delta_{i,0,t}I\{C_{i,t-1} = 0\}$.

Since the random terms $\varepsilon_{i,j,t}$ are i.i.d, the conditional probability of the entire choice sequence $C_i = C_{i,1}\ldots C_{i,T}$ is equal to the product of conditional choice probabilities:

$$\begin{aligned}&\Pr\left(C_i \mid X, Y_i, N_i^a, N_i^e, \omega_i, \nu_i, \upsilon_k, \theta\right)\\&\quad=\prod_{t=1}^{T}\Pr\left(C_{i,t}=j \mid C_{i,t-1}, x_t, Y_i, N_{i,t}^a, N_{i,t}^e, \omega_{i,t}, \nu_{i,t}, \upsilon_k, \theta\right)\end{aligned} \tag{17}$$

where $X = \{x_1, \ldots, x_T\}$, $\omega_i = \{\omega_{i,1}, \ldots, \omega_{i,T}\}$, $\nu_i = \{\nu_{i,1}, \ldots, \nu_{i,T}\}$, $N_i^a = \{N_{i,1}^a, \ldots, N_{i,T}^a\}$ and $N_i^e = \{N_{i,1}^e, \ldots, N_{i,T}^e\}$.

The history probability depends on the unobservables $\omega_i, \nu_i, \upsilon_k$. Integrating them out yields:

$$\Pr\left(C_i \mid X, Y_i, N_i^a, N_i^e, \Omega\right)=\sum_{k=1}^{K}\int \Pr\left(C_i \mid X, Y_i, N_i^a, N_i^e, \tilde{\omega}, \tilde{\nu}, \upsilon_k, \theta\right)dF(\tilde{\omega}, \tilde{\nu})p_k \tag{18}$$

where $F(\tilde{\omega}, \tilde{\nu})$ is the joint c.d.f. of ω, ν, and Ω includes all the parameters of θ as well as the segment-specific parameters $\upsilon_1, \ldots, \upsilon_K$ and the parameters of the density functions of ω, ν and υ.

Since all the unobservables are independent across individuals, the log-likelihood is simply

$$\ln L(\Omega)=\sum_{i=1}^{I}\ln\left[\Pr\left(C_i \mid X, Y_i, N_i^a, N_i^e, \Omega\right)\right] \tag{19}$$

where I is the number of consumers in the sample.

The computation of the likelihood function involves integration over the joint distribution of the signals and the show random effects. Since numerical integration by quadrature techniques is infeasible, we resort to Monte-Carlo integration.

4.3. Identification

The identification of this model under risk neutrality is discussed at length in Anand and Shachar (2001). Thus, we focus here on the identification of the risk parameter. Furthermore, we show how the risk-reduction effect of advertising can be distinguished from the direct effect.

The risk parameter ϕ is identified by many moments. This parameter multiplies $\sigma_{i,j,t}$ which is a function of all the characteristics of individuals and the attributes of shows. Indeed, these variables also affect $\mu_{i,j,t}^p$ (and thus assist in identifying additional parameters in the model), however, their effect on $\sigma_{i,j,t}^2$ is quite different than their influence on $\mu_{i,j,t}^p$. Specifically, while $\mu_{i,j,t}^p$ is based on the *mean* of the show attributes (for each network), $\sigma_{i,j,t}^2$ is based on the *variance-covariance matrix* of the show attributes.

Advertising increases the purchase probability of risk-averse consumers for two reasons: (1) it reduces the risk associated with the product, and (2) it directly increases the utility through the ρ function. The distinction between these two sources is based on equation (10). Specifically, while the effect through the ρ function does not depend on factors such as the consumer's familiarity with the product and involvement level, the risk-reduction effect does. For example, if the positive effect of $N^a_{i,j,t}$ on the purchase probability is fully explained by the interaction of the risk parameter and $\sigma^2_{i,j,t}$, then we would detect no additional effect of $N^a_{i,j,t}$ through the ρ function. However, if this positive effect cannot be fully explained by the risk parameter, the parameters of the ρ function would not be zero.[22]

4.4. The endogeneity problem

The endogeneity problem has accompanied empirical research of advertising since its inception. In individual-level data, the primary source of the problem comes from the targeting strategies of firms. Firms tend to send ads to consumers who, a priori, have a higher tendency to consume the promoted product. Thus, the tendency of consumers to purchase a product is positively correlated with their exposure to advertising even if their expected utility is not affected in any way by ads. If consumers' a priori tendencies are unobserved by the researcher, there is an endogeneity problem—an observed explanatory variable (ad exposures) is correlated with the unobservables of the model. Thus, targeting leads to a bias in the estimate of ad effectiveness.

In general, in order to resolve the endogeneity problem, one needs to endogenize $N^a_{i,j,t}$. Such a solution requires that the researcher model both the consumption decision and the exposure to advertising, and collect data on these two parts of the model. Our empirical application makes this difficult task easier and feasible.[23] Specifically, it is easy to show that the endogeneity of advertising in our setup is equivalent to the standard state dependence/heterogeneity problem. To see that, notice that one can rewrite the exposure to advertising in our model as: $N^a_{i,j,t} = \sum_{\tau=1}^{t-1} a_{j,t,\tau} I\{C_{i,\tau} = j\}$, where $a_{j,t,\tau}$ is a binary variable equal to 1 if an ad for a show on network j at time t was aired in time slot τ. Notice that $a_{j,t,\tau}$ is not individual specific, and what makes the ad exposure, $N^a_{i,j,t}$, individual specific are her viewing choices. Thus, in a sense, the ad exposure variable is a stock of some of the previous viewing choices. This is equivalent to the definition of state dependence (see, for example, pp. 139–140 in Heckman, 1981). This means that the solution to the endogeneity of advertising in our setup is the same as the solution to the state dependence challenge in general discrete panel data models—it rests on properly accounting for the unobserved heterogeneity parameters. The advantage of our dataset is that the consumption decision also determines the exposure to advertising. Thus, within the same framework and dataset, we model both the consumption decision and the exposure to advertising.

Monte Carlo experiments support the suggested solution for the endogeneity problem in our setup. Specifically, we find that the estimation procedure identifies correctly the number of segments in the population and, as a result, the estimates of advertising effectiveness are consistent.

22 Thus, the direct effect on the utility can be also considered as a residual (i.e., the excess effect of advertising on choices that is not captured by the other elements of the model).

23 Previous studies of advertising that used individual level data (other than Shachar and Anand, 1998; Anand and Shachar, 2001) have not endogenized the exposure to advertising variable.

5. The results

We estimate various specifications of the model. We refer to the specification presented in the previous sections as the "baseline model" and denote it in the tables as Model I. In Model II we impose risk-neutrality ($\phi = 0$). The differences between the estimates of Models I and II demonstrate that ignoring the risk-reduction role of advertising might lead to misleading managerial implications. The estimates of the other models (Models III-VIII) show that our results are robust to various specifications of the model.[24]

As discussed above, the risk-reduction role of advertising stems from (a) consumers' uncertainty (i.e, $(\varsigma^m_{i,j})^{-1} > 0$), (b) consumers' risk-aversion (i.e., $\phi < 0$) and (3) the informative content of advertising (i.e, $\varsigma^a > 0$). In all the specifications the risk-reduction role of advertising is supported by the data (i.e., $(\widehat{\varsigma^m_{i,j}})^{-1} > 0$, $\hat{\phi} < 0$, and $\widehat{\varsigma^a} > 0$).

The estimates of the other parameters are specific to the television example, and therefore are of less interest. However, we briefly report them as well (for the baseline specification only), as a reality check on the model.

The estimates are presented in Tables 2 to 9.[25] The number of latent segments was determined by minimizing the BIC (Bayes Information Criterion). For the baseline model, the BIC was minimized at 3 segments. Segments' sizes are 0.23, 0.27 and 0.50.

5.1. Estimates

5.1.1. Utility parameters. *Show attributes (βs and νs: Table 2):* The β estimates are consistent with the findings of previous studies. Viewers prefer shows whose cast demographics are similar to their own. Viewers differ in both observed and unobserved ways in their taste for particular show genres. For example, the utility from sitcoms decreases with age.

State dependence parameters (δ: Table 3): As in previous studies, state dependence *between* shows is an important factor. The estimates of υ^δ_i range from 1.53 to 2.32. There is obviously also state-dependence *during* a show. The state dependence during a show depends on the show type. It is weaker for the News magazines and Sports shows than for the other show categories ($\delta_{News,Sports} = -0.51$, $\delta_{Sitcom} = 0.00$, $\delta_{AD} = 0.14$, $\delta_{RD} = 0.15$). This could be explained by lack of a continuous script and plot in news magazines and sports shows.[26] Most of the demographic variables are insignificant, except for $\delta_{Female} = 0.13$.

24 The estimation was done in Gauss. All the computationally-intensive parts of the model were written in C and packaged into a DLL (dynamic link library) accessible from Gauss. Since the simulator we use is smooth, standard gradient-based optimization methods can be used. Hajivassiliou (1997) suggests increasing the number of simulation draws R until the expectation of the score function is zero at $\hat{\Omega}'_{MSL}$. In our case this is achieved at $R = 400$. Each iteration takes about 10 minutes on a Pentium-IV 2.8 GHz computer.

25 The standard errors were computed from the inverse of the information matrix, thus they ignore the additional error induced by the simulation noise.

26 Compared with the state dependence between shows, the state dependence during a show is weaker for News and Sports shows and stronger for both types of drama. However, this does not imply that the viewing persistence during News and Sports shows is lower than the persistence between shows. Note that other elements of the model (e.g. show random effects) also generate persistence in choices during the show, but not between shows, thus the total persistence in choices during News and Sports shows is higher than between show.

Table 2. Preferences for show attributes (β).

Parameter	Estimate	Parameter	Estimate
β_{Gender}	0.2036 (0.0440)	$\beta_{AD}^{Education}$	−0.3253 (0.2233)
β_{Age0}	0 (–)	β_{AD}^{Family}	−0.0110 (0.1163)
β_{Age1}	−0.2155 (0.0459)	β_{AD}^{Family}	−0.2450 (0.0987)
β_{Age2}	−0.7081 (0.0757)	β_{RD}^{Teens}	0 (–)
β_{Family}	0.4160 (0.0970)	β_{RD}^{GenX}	0.0307 (0.2462)
β_{Race}	−1.0019 (0.2184)	$\beta_{RD}^{BabyBoomer}$	−0.1016 (0.2351)
β_{Sitcom}^{Teens}	0 (–)	β_{RD}^{Older}	−0.6952 (0.2572)
β_{Sitcom}^{GenX}	−0.7109 (0.2009)	β_{RD}^{Female}	0.5653 (0.1170)
$\beta_{Sitcom}^{BabyBoomer}$	−0.7475 (0.1971)	β_{RD}^{Income}	−1.7525 (0.2722)
β_{Sitcom}^{Older}	−1.3964 (0.2105)	$\beta_{RD}^{Education}$	−0.2904 (0.2554)
β_{Sitcom}^{Female}	0.4346 (0.0882)	β_{RD}^{Family}	−0.0423 (0.1295)
β_{Sitcom}^{Income}	−0.2310 (0.2075)	β_{RD}^{Urban}	−0.0728 (0.1112)
$\beta_{Sitcom}^{Education}$	−0.3443 (0.2017)	$\upsilon_{k=1}^{Sitcom}$	0 (–)
β_{Sitcom}^{Family}	0.1481 (0.1191)	$\upsilon_{k=1}^{AD}$	0 (–)
β_{Sitcom}^{Urban}	−0.0186 (0.0882)	$\upsilon_{k=1}^{RD}$	0 (–)
β_{AD}^{Teens}	0 (–)	$\upsilon_{k=2}^{Sitcom}$	−0.9028 (0.1488)
β_{AD}^{GenX}	−0.6748 (0.2156)	$\upsilon_{k=2}^{AD}$	−0.2861 (0.1666)
$\beta_{AD}^{BabyBoomer}$	−0.3622 (0.2089)	$\upsilon_{k=2}^{RD}$	0.5721 (0.1874)
β_{AD}^{Older}	−0.2703 (0.2180)	$\upsilon_{k=3}^{Sitcom}$	−0.8561 (0.1524)
β_{AD}^{Female}	0.3716 (0.0974)	$\upsilon_{k=3}^{AD}$	−0.0874 (0.1636)
β_{AD}^{Income}	−0.8713 (0.2473)	$\upsilon_{k=3}^{RD}$	−0.2001 (0.1890)

Notes on normalizations: we set $\beta_{Teens}^{Sitcom} = \beta_{Teens}^{AD} = \beta_{Teens}^{RD} = \upsilon_{k=1}^{Sitcom} = \upsilon_{k=1}^{AD} = \upsilon_{k=1}^{RD} = 0$ because we estimate a fixed effect parameter for each of the shows. Standard errors are in parentheses.

Table 3. State-dependence parameters.

Parameter	Estimate	Parameter	Estimate
$\upsilon^{\delta}_{k=1}$	2.2832 (0.0978)	δ_{GenX}	−0.0989 (0.0692)
$\upsilon^{\delta}_{k=2}$	1.5251 (0.0985)	δ_{Boomer}	−0.0432 (0.0663)
$\upsilon^{\delta}_{k=3}$	2.3221 (0.1101)	δ_{Old}	0.0216 (0.0655)
δ_{Out}	0.2937 (0.0942)	δ_{Female}	0.1346 (0.0392)
$\delta_{News,Sports}$	−0.5096 (0.1564)	$\delta_{Hour(out)}$	−0.4322 (0.0821)
δ_{AD}	0.1354 (0.1617)	δ_{Last15}	0.1953 (0.1564)
δ_{RD}	0.1472 (0.1622)	$\delta_{First15}$	0.1576 (0.1038)
δ_{Sitcom}	0.0032 (0.1611)	$\delta_{InProgress}$	−0.3557 (0.0655)
δ_{Teen}	0 (–)		

Note on normalizations: we set $\delta_{Teens} = 0$ because we estimate υ^{δ}_{k} for all the segments. Standard errors are in parentheses.

Table 4. Outside alternative parameters.

Parameter	Estimate	Parameter	Estimate
$\gamma_{Premium}$	0.4471 (0.0543)	γ_{Female}	0.0974 (0.0934)
γ_{Basic}	0.3010 (0.0515)	γ_{Income}	−0.4507 (0.2089)
γ_{Teens}	−3.4099 (0.4820)	$\gamma_{Education}$	−0.0976 (0.2027)
γ_{GenX}	−4.1281 (0.5188)	γ_{Family}	−0.0371 (0.1089)
γ_{Boomer}	−4.1958 (0.5178)	γ_{Urban}	−0.1006 (0.0903)
γ_{Old}	−4.7663 (0.5094)		

Note: Standard errors are in parentheses.

Finally, the viewers do not like switching into the middle of another show ($\delta_{InProgress} = -0.36$).

Preference for the outside option (γ: Table 4): The utility from the outside alternative decreases with age and income. It is the lowest for viewers who do not have a cable connection, and the highest for those who have a premium subscription. Recall that the outside option lumps together two alternatives: TV off and non-network channels. All the other demographic variables do not have a significant effect on the utility from the outside alternative.

The outside alternative is also characterized by substantial state-dependence. The estimate of $\upsilon^{\delta}_{i} + \delta_{Out}$ ranges from 1.82 to 2.62. As expected, the state-dependence in the outside alternative declines "on the hour" ($\delta_{Hour} = -0.43$), when many non-network channels start airing new shows.

Fixed and random effects (η, α and σ_{ν}: Tables 5 and 6): Table 5 presents the fixed effects for each show. We observe substantial differences in these parameters. The estimates range from −1.10 for Dateline NBC (Friday) to 2.59 for the X-Files.

Table 6 presents the distribution parameters of the random individual-network effects. There is a large heterogeneity in these parameters. For example, the second and the third

Table 5. Show-specific fixed effects.

Show	Estimate	Show	Estimate	Show	Estimate	Show	Estimate
ABC		NBC		CBS		Fox	
The Marshal	0	Fresh Prince of Bel-Air	2.1229	The Nanny	1.7090	Melrose Place	2.0941
	(–)		(0.4128)		(0.4101)		(0.4416)
Pro Football	−0.0237	In the House	1.6492	Can't Hurry Love	1.7170	Beverly Hills 90210 (Mon)	2.0340
	(0.3603)		(0.4229)		(0.3953)		(0.4512)
Roseanne	0.5478	She Fought Alone	0.6006	Murphy Brown	1.4211	Bram Stoker's Dracula	1.2471
	(0.2844)		(0.4070)		(0.3999)		(0.3471)
Hudson Street	0.7774	Wings	1.7870	High Society	1.2462	Beverly Hills 90210 (Wed)	1.9473
	(0.3063)		(0.3854)		(0.4070)		(0.4368)
Home Improvement	1.5306	News Radio	1.3785	The Client	0.3587	Party of Five	0.7856
	(0.2954)		(0.3898)		(0.2965)		(0.4345)
Coach	0.9071	Frasier	1.8917	Nothing Lasts Forever	1.2899	Living Single	2.0954
	(0.2985)		(0.3938)		(0.4183)		(0.4425)
Ellen	0.9119	Pursuit of Happiness	0.2605	Bless this House	0.4824	The Crew	2.4840
	(0.2821)		(0.3800)		(0.3907)		(0.4917)
The Drew Carey Show	0.6474	Seaquest 2032	0.3332	Dave's World	1.1809	New York Undercover	1.0869
	(0.3039)		(0.2932)		(0.4094)		(0.4645)
Grace Under Fire	1.4044	Dateline NBC (Wed)	−0.4830	Central Park West	0.4990	Strange Luck	1.2268
	(0.2941)		(0.4105)		(0.4025)		(0.3570)
The Naked Truth	0.6244	Friends	2.1445	Murder, She Wrote	0.0686	X-Files	2.5893
	(0.3034)		(0.3781)		(0.2913)		(0.3637)
Columbo	−0.3959	The Single Guy	1.3924	New York News	−0.4092		
	(0.1649)		(0.3838)		(0.3046)		
Family Matters	1.6717	Seinfeld	2.0116	Here Comes the Bride	0.4451		
	(0.3146)		(0.3785)		(0.3851)		
Boy Meets World	0.5199	Caroline in the City	1.0364	Ice Wars	−0.8955		
	(0.3236)		(0.3734)		(0.4137)		
Step by Step	0.1777	Unsolved Mysteries	0.5188				
	(0.3084)		(0.2838)				
Hanging with Mr. Cooper	1.0262	Dateline NBC (Fri)	−1.0989				
	(0.3384)		(0.3988)				

Notes on normalization: we set the fixed effect of the show *The Marshal* to zero because we estimate the fixed effect of the outside alternative. Standard errors are in parentheses.

Table 6. Random individual-network effects.

Parameter	Estimate	Parameter	Estimate
σ_ν	0.5895 (0.0940)	μ_2	0.1555 (0.1578)
$\alpha_{8:00,\text{ABC}}$	−0.9041 (0.1029)	$\alpha_{k=2,\text{ABC}}$	−0.8266 (0.3579)
$\alpha_{8:00,\text{CBS}}$	−0.4542 (0.1072)	$\alpha_{k=2,\text{CBS}}$	−0.8066 (0.2854)
$\alpha_{8:00,\text{NBC}}$	−0.8866 (0.1019)	$\alpha_{k=2,\text{NBC}}$	−0.7876 (0.3015)
$\alpha_{8:00,\text{FOX}}$	−0.2659 (0.1041)	$\alpha_{k=2,\text{FOX}}$	−1.1393 (0.3344)
μ_1	0 (–)	μ_3	0.7823 (0.1085)
$\alpha_{k=1,\text{ABC}}$	0 (–)	$\alpha_{k=3,\text{ABC}}$	−1.5879 (0.3392)
$\alpha_{k=1,\text{CBS}}$	0 (–)	$\alpha_{k=3,\text{CBS}}$	−1.1060 (0.3276)
$\alpha_{k=1,\text{NBC}}$	0 (–)	$\alpha_{k=3,\text{NBC}}$	−1.4829 (0.2999)
$\alpha_{k=1,\text{FOX}}$	0 (–)	$\alpha_{k=3,\text{FOX}}$	−2.1601 (0.3408)

Note: we set the random effect parameters for the first segment at zero, because we estimate all the fixed effects of shows. Standard errors are in parentheses.

segments like FOX less than other networks, and the third segment prefers CBS over other networks.

In addition, we find a significant show random effect ($\sigma_\nu = 0.59$, $s.e. = 0.09$). Recall that the α-s capture the unobserved individual-network match, while σ_ν captures the unobserved individual-show match.

The direct effect parameters (ρ: Table 9a, Model I): The estimated direct effect is close to zero for all three segments. It is positive for the second and third segments, and negative for the first segment. Although the coefficients are statistically significant, we show in Section 5.2 that the direct effect of advertising is behaviorally negligible.

5.1.2. Information set parameters (ς: Tables 7 and 8). In our model, the individual obtains information from the brand images of the firms, $\mu_{i,j}$, and from three kinds of unbiased noisy signals. While the precision of the brand image, $\varsigma_{i,j}^{p}$, is determined by the variance of attribute utility, the precisions of the signals are free parameters in estimation.

The key parameter of interest is the precision of the advertising signals (Table 8, Model I). The estimate of $\sqrt{\varsigma^a}$ is 1.55, with a standard error of 0.18, implying that advertising content is informative and that consumers update their expectation based on the information in the

Table 7. Precision of miscellaneous signals.

Parameter	Segment 1	Segment 2	Segment 3
$\sqrt{\varsigma^m_{k=1,j=\text{ABC}}}$	3.0281 (0.3966)	3.3257 (0.4682)	3.3904 (0.4698)
$\sqrt{\varsigma^m_{k=1,j=\text{CBS}}}$	2.6509 (0.3767)	2.0712 (0.3568)	2.1606 (0.3650)
$\sqrt{\varsigma^m_{k=1,j=\text{NBC}}}$	2.7919 (0.3853)	2.9095 (0.4106)	2.7203 (0.4170)
$\sqrt{\varsigma^m_{k=1,j=\text{FOX}}}$	1.7660 (0.2993)	2.4329 (0.3485)	2.3928 (0.3443)

Note: Standard errors are in parentheses.

Table 8. Risk-aversion and information parameters.

Parameter	Model I	Model II	Model III	Model IV	Model V	Model VI	Model VII	Model VIII
ϕ	−15.8410	0	−15.8580	−15.7509	–	−15.9111	−15.2589	−15.9162
	(2.4226)	(–)	(4.4635)	(2.3382)		(2.8568)	(2.7489)	(2.3595)
$\phi_{k=1}$	–	–	–	–	−16.7828	–	–	–
					(3.3359)			
$\phi_{k=2}$	–	–	–	–	−17.0482	–	–	–
					(4.0067)			
$\phi_{k=3}$	–	–	–	–	−15.0262	–	–	–
					(2.7494)			
$\sqrt{\varsigma^a}$ (Mon–Tue)	1.2612	0.0280	1.1834	0	1.0411	1.1803	1.5652	1.2192
	(0.1826)	(2.2521)	(0.1888)	(–)	(0.1396)	(0.2325)	(0.2749)	(0.1793)
$\sqrt{\varsigma^a}$ (Wed–Fri)	1.5503	1.1434	1.4843	0	1.3501	1.5046	1.7464	1.4981
	(0.1823)	(0.3850)	(0.1786)	(–)	(0.1577)	(0.2290)	(0.2645)	(0.1824)
$\sqrt{\varsigma^e}$	3.1528	0.0174	0	3.2260	3.0200	3.1721	3.3245	3.1732
	(0.4251)	(0.6658)	(–)	(0.4027)	(0.4087)	(0.4297)	(0.4652)	(0.4282)
$\sqrt{\varsigma^m_{j=\text{ABC}}}$ *	3.2898	2.4253	3.3044	3.2477	3.1470	3.2874	3.2449	3.2820
$\sqrt{\varsigma^m_{j=\text{CBS}}}$ *	2.2492	0.6693	2.1390	2.2470	2.1556	2.2468	2.1891	2.2425
$\sqrt{\varsigma^m_{j=\text{NBC}}}$ *	2.7875	2.0826	2.8394	2.8369	2.6434	2.7825	2.6996	2.7806
$\sqrt{\varsigma^m_{j=\text{FOX}}}$ *	2.2596	0.6328	1.9853	2.3116	2.1455	2.2510	2.2433	2.2576

* Weighted average of $\sqrt{\varsigma^m_{k,j}}$ across segments.
Model I – the baseline specification.
Model II – the risk-neutral model ($\phi = 0$).
Model III – experience signals are disabled ($\varsigma^e = 0$).
Model IV – all the effects of advertising are disabled ($\varsigma^a = 0,\ \upsilon^\rho_k = 0$).
Model V – segment-specific risk parameter (ϕ_k instead of ϕ)
Model VI – a non-parametric specification of the persuasive effect:

$$\rho_k(N) = \rho_{1,t,k} I\{N = 1\} + \rho_{2,t,k} I\{N = 2\} + \rho_{3,t,k} I\{N = 3\} + \rho_{4,t,k} I\{N \geq 4\}$$

where $\rho_{n,t,k} = \rho_{n,MT,k}\text{MT}_t + \rho_{n,WF,k}\text{WF}_t$
Model VII – a richer parametric specification of the persuasive effect:

$$\rho_k(N) = \left(\rho_{1,t,k} + \rho^Y_1 Y_i\right)N + \left(\rho_{2,t,k} + \rho^Y_2 Y_i\right)N^2$$

where $\rho_{1,t,k} = \rho_{1,MT,k}\text{MT}_t + \rho_{1,WF,k}\text{WF}_t,\ \rho_{2,t,k} = \rho_{2,MT,k}\text{MT}_t + \rho_{2,WF,k}\text{WF}_t$.
Model VIII – the persuasive effect is specified as:

$$\rho_k(N) = (\rho_{\text{MT},k}\text{MT}_t + \rho_{\text{WF},k} WF_t) I\{N > 0\}.$$

Standard errors are in parentheses.

ads.[27] The informative content of ads is one of the building-blocks in the risk-reduction role of advertising. The impact of this effect is discussed in Section 5.2.

The estimate of $\sqrt{\varsigma^e}$ is 3.15, with a standard error of 0.43. Thus, the experience signals are informative. Furthermore, as one might expect, the information acquired after watching any 15 minutes of the show is more precise than the information from one ad. Specifically, the information acquired from two ads is slightly less than the information acquired after watching any 15 minutes of the show.

The precision of the miscellaneous signals differs across segments and networks (Table 7). The point estimates of $\sqrt{\varsigma^m_{i,j}}$ range from 1.77 for FOX in the first segment, and up to 3.39 for ABC in the third segment. On average, the network with the best known shows is ABC, and the networks with the least known programs are CBS and FOX (Table 8, Model I). These estimates are sensible for the following reasons. The degree of familiarity with a network should be an increasing function of (a) the ratings of its shows and (b) the "age" of its shows (i.e., the number of seasons that the shows were on the air). Even though NBC enjoyed the highest average rating during the fall season of 1995 (followed by ABC in second place), it was only third in the "ratings race" during the 1994 season (behind ABC and CBS). Moreover, while several of NBC's highest rated shows in 1995 were in their first year of airing, the successful ABC shows were veterans. Thus, the finding that the network with the best known shows is ABC seems reasonable. The low $\hat{\varsigma}^m_{i,j}$ 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.

A test of the uncertainty assumption. We have also estimated a model with the following restriction $(\varsigma^m_{i,j})^{-1} = 0$ for all $i,\ j$. This model assumes that consumers are fully informed about product attributes. The perfect information hypothesis is easily rejected by the likelihood ratio test (the statistical χ^2 is 447.8, and the critical, at the one percent level, is 26.2). Consumers' uncertainty about products' attributes is another building-block in the risk-reduction role of advertising.

5.1.3. Risk aversion parameter (ϕ: Table 8, Model I). The estimate of the risk-aversion parameter ϕ is negative and significant both behaviorally and statistically ($\phi = -15.84$, with a standard error of 2.42), implying that consumers are clearly risk-averse. Consumers' risk aversion is the final building-block in the risk-reduction role of advertising. Next, we examine the risk-reduction effect.

5.2. *The risk reduction role of advertising (empirical)*

The results show that in our data (a) consumers are uncertain about products' attributes (i.e., $(\widehat{\varsigma^m_{i,j}})^{-1} > 0$), (b) they are risk-averse (i.e., $\hat{\phi} < 0$) and (3) the ad content is informative (i.e., $\widehat{\varsigma^a} > 0$). In our model these findings imply that advertising has a risk-reduction role. In other words, exposure to advertising increases consumers' tendency to purchase

27 Due to missing data in exposures to ads in the first days of our dataset, we estimate two separate parameters for Monday—Tuesday and Wednesday-Friday. The estimates reported pertain to Wednesday—Friday.

the promoted product *because* advertising provides information that resolves some of the uncertainty that the risk-averse consumers face and thus reduces the risk associated with the promoted product.

The following exercise assesses the magnitude of this risk-reduction effect. First, we set the persuasive effect of ads to zero for all the shows (in order to isolate the risk-reduction effect). Then, for each show on each network, we expose the individuals in our data to 0, 1, 2 or 3 ads, and compute the predicted market share of that show.[28] The average increase in the market share is 2.7 percentage points (an increase of 53.3%) after the first exposure, 5.4 percentage points (108.4%) after 2 exposures (compared to 0 exposures), and 8.0 percentage points (162.3%) after 3 advertising exposures.[29]

Our model allows advertising to affect choices also directly via the ρ function. It is interesting to compare the effectiveness of advertising through the risk-reduction channel with the standard effect of advertising via the ρ function. Thus, we repeat the above exercise in order to assess the magnitude of the direct effect. Specifically, for each show on each network, we set the ad exposure to 0, 1, 2 or 3 in the direct effect, but retain the original ad exposure in the risk-reduction effect. This procedure allows us to measure the magnitude of the direct effect separately from the risk-reduction effect. On average, the predicted market share of a show increases by just 0.08 percentage points (1.5%) after the first exposure, and by 0.18 percentage points (3.1%) after 2 or 3 advertising exposures (compared to 0 exposures).

This means that in our data, not only that the risk-reduction effect is strong and significant, but it is also much more important than the standard effect of advertising. In other words, the direct effect of advertising is much smaller than the indirect effect, and it is negligible behaviorally. Note that the direct effect is identified as a reduced-form residual effect of ad exposures on choices, after controlling for risk-reduction. This residual effect could actually be capturing various other effects of ads. However, our finding that the direct effect of ads is negligible resolves this interpretation problem in our data, and suggests that the structure of our model is rich enough to capture consumer behavior.

The conclusion from these findings is that in our data advertising has a negligible direct effect on the utility. In other words, exposure to advertising does not change the preferences of consumers. The positive effect of advertising on the purchase probability, which is usually interpreted as an indication for the persuasive power of advertising, is actually due to the risk-reduction effect of ads. We discuss the welfare implications of this result in the concluding section.

28 We hold the ad exposures for all the other shows fixed at their actual levels. For each show and number of exposures, we simulate 100 choice sequences for every individual to compute the market share of the show. We repeat this procedure for every show on Wednesday, Thursday and Friday, from 0 to 3 advertising exposures.

29 Note that on average, each ad aired by the networks for Wednesday-Friday shows generates just 0.08 advertising exposures, thus the high advertising effectiveness *per exposure* is quite plausible. When we conduct a related experiment in which we *air* an additional ad (with exposures determined by viewing choices), the average increase in the market share is just 0.19 percentage points (an increase of 3.9%). Notice that since at the individual level both the risk-reduction effect and the direct effect depend on the number of advertising *exposures*, it's more natural to express all the behavioral implications of the model in terms of exposures, as opposed to ads aired.

It is worth noting that Erdem and Keane (1996), who also found that advertising content is informative and consumers are risk-averse, did not allow advertising to have a direct effect on the utility. Thus, the only channel through which ad exposure can have a positive effect on the purchase probability in their model is the risk-reduction effect. Our model includes both avenues (risk-reduction and direct effect). This has two benefits: (a) the finding that advertising has a risk-reduction effect is not due to mis-specification of the utility, and (b) we can conclude that there is no evidence for the direct effect of advertising.[30]

5.3. The implications of risk-reduction

The risk-reduction role of advertising has various implications. For example, unlike the traditional models (which only include the direct effect), our model suggests that the effectiveness of advertising depends on the familiarity of consumers with the promoted products. Indeed, this implication is shared with other models of informative advertising. However, other implications are quite unique. For example, our model suggests that advertising effectiveness should be higher for the high-involvement consumers than for the low-involvement consumers. Here, we demonstrate this implication and in the concluding section we discuss its managerial consequences.

In order to illustrate this effect, we need to construct a measure of consumers' involvement. Our measure of involvement for consumer i is $Involvement_i$. This variable is equal to $\sum_k P_i^k \mathrm{Var}(x_{j,t}\beta_i^k + \eta_{j,t})$, where P_i^k is the posterior probability that i is of segment k, and β_i^k is her taste parameter in that case.[31] The variance is computed over all networks and timeslots. Thus, this measure captures how much the individual cares (in terms of attribute utility) about the differences between the products she is facing. Individuals with high βs (in absolute terms) are more sensitive to products attributes. For them $Involvement_i$ would be relatively high.

We illustrate the effect of involvement on advertising effectiveness by the following experiment. We split all the individuals into 3 groups of equal size according to their $Involvement_i$. For each group, we simulate the predicted market share for each show, with 0 and 1 ad exposures.[32] On average, exposure to 1 ad increases the market share by 2.47 percentage points for the lowest-involvement group, by 2.52 percentage points for the next group, and by 3.22 percentage points for the highest-involvement group.[33] Thus, our estimates suggest that the advertising effectiveness is substantially higher for

30 Furthermore, EK did not account for the endogeneity of the ad exposure variables, and, as discussed in Anand and Shachar (2001), their identification rests entirely on the structure that the model imposes on the variance-covariance matrix of the random components. Our observable characteristics introduce additional identifying moments and are likely to improve the robustness of the results. Note that the focus of EK is not on advertising effectiveness, and thus their findings on the role of advertising are not central in their study.

31 Recall that β_i varies both with the segments and with demographics y_i.

32 For each show and number of exposures, we simulate 100 choice sequences (using posterior segment probabilities) for all the individuals to compute the market share of the show. The ad exposures for all the other shows are set to zero. The direct effect is set to zero.

33 The difference between the highest-involvement group and the medium-involvement group is significant at 1%.

the high-involvement individuals, and reinforce the importance of proper targeting of ads.[34]

5.4. A comparison to the standard advertising model

The risk-reduction effect has been ignored in the standard advertising model. Our findings indicate that the standard model of advertising is mis-specified. Such mis-specification might lead to inconsistent estimates and misleading implications. Here, we illustrate these consequences of mis-specification.

For this purpose we re-estimate the model without the risk-reduction effect by imposing risk-neutrality ($\phi = 0$, Tables 8 and 9a, Model II). In this case, the estimated direct effect of ads is positive and highly significant. We repeat the same experiment to measure the magnitude of the direct effect in the risk-neutral model. On average, the direct effect of ads increases the predicted market share of a show by 2.6 percentage points (50.2%) after 1 exposure, by 6.0 percentage points (115.9%) after 2 exposures, and by 10.0 percentage points (194.0%) after 3 advertising exposures.

This means that when the risk-reduction is ignored, we find, like previous studies of advertising, that ads have a strong direct effect on the utility. This finding is an immediate result of the mis-specification of the model. In other words, it is likely that the findings of previous studies that ads are persuasive are misleading.

Furthermore, by ignoring the risk-reduction effect and attributing the positive correlation between ad exposure and the tendency to consume the promoted product to the direct effect, previous studies fail to identify and realize the richness of advertising effectiveness. An important ingredient in any advertising strategy is to identify the consumers and products for which ads are especially effective. The standard advertising model does not account for the variation in ad effectiveness as a result of variation in (a) the involvement level, (b) the diversity of products offered by the firms, (c) the risk parameter, (d) the precision of ads' signals, (e) the familiarity of consumers with products.[35]

Table 9a. Persuasion parameters (Models I, II, III, V).

Parameter	Model I	Model II	Model III	Model V
ρ_1 (Mon–Tue)	−0.2874 (0.1210)	0.7485 (0.1166)	−0.3053 (0.1628)	−0.1650 (0.1164)
ρ_2 (Mon–Tue)	0.0570 (0.0297)	−0.1374 (0.0314)	0.0654 (0.0364)	0.0430 (0.0306)
ρ_1 (Wed–Fri)	−0.2556 (0.0927)	0.5912 (0.0804)	−0.2491 (0.1322)	−0.0842 (0.0587)
ρ_2 (Wed–Fri)	0.0512 (0.0204)	−0.026 (0.0150)	0.0582 (0.0266)	0.0114 (0.0103)
$\upsilon^{\rho}_{k=1}$	1 (–)	1 (–)	1 (–)	1 (–)
$\upsilon^{\rho}_{k=2}$	−0.4948 (0.4718)	0.6897 (0.1090)	−0.6011 (0.7059)	−2.6023 (2.0558)
$\upsilon^{\rho}_{k=3}$	−1.0963 (0.6896)	1.1874 (0.1666)	−1.2333 (1.0554)	−3.1336 (2.5061)

Note: Standard errors are in parentheses.

34 In the working paper version of this study, we conduct a similar exercise with respect to the familiarity level (instead of the involvement level) and find that ad effectiveness is a decreasing function of familiarity.

35 Note that while the standard model does not account for the effect of familiarity on the effectiveness of advertising, informative advertising models do.

Table 10 illustrates the differences between the implications of the standard model and the risk-reduction model. It compares the predicted ad effectiveness between the two models. We find that the standard model overstates the average advertising effectiveness for ABC and NBC (2.5 percentage points versus 1.8, and 3.1 versus 2.8), and understates advertising effectiveness for CBS and FOX (1.8 versus 2.4, and 2.6 versus 3.3).[36] Note that this direction of bias in the standard model is consistent with the findings in Section 5.2, that ABC and NBC are the best-known networks while CBS and FOX are the least-known. Furthermore, for some shows the predictions of the standard model display a very substantial bias (for example, 1.5 versus 4.1 percentage points per exposure for FOX's show "Strange Luck"). Thus, in our empirical setting, the advertising strategy prescribed by the standard model is likely to be quite different quantitatively from the optimal strategy.

5.5. A comparison to the signaling models of advertising

In some aspects the signaling theory of advertising is closer to our approach than the standard model of advertising. Both theories suggests that exposure to ads can increase the tendency to purchase the promoted product even if the utility is not a function of ads. Furthermore, both imply that ad effectiveness is a decreasing function of consumers' knowledge about the promoted product.

However, as we discuss above, the implications of the risk-reduction model (and thus the identification of the model) are richer than those of the signaling theory. Specifically, ad effectiveness in the risk-reduction model is determined by both familiarity (like in signaling) and other factors (specific to the risk-reduction model).

While the standard model of advertising is nested in our model, the signaling theory of advertising is not. We have not integrated the signaling approach into our model for various reasons. The main reason is that the risk-reduction model and the signaling theory focus on different types of products and ads. While the first applies to search goods, the second is relevant for experience goods. (An individual can know her utility from a search good, but not from an experience good, even without consuming it. See Tirole, 1988, p. 106). For search goods, ads content can credibly convey information on products' attributes (i.e., "this automobile has 200 horse powers"). However, for experience goods, experience is required in order to acquire information about the utility, and ads content cannot credibly convey information on product quality. Thus, according to Nelson (1974), ad intensity (not ad content) serves as a credible signal about the quality of the product. This means that the two models deal with different types of products and different types of ads. Moreover, the following exercise suggests that the signaling theory is probably not relevant for our empirical example.

According to the signaling model, ad intensity should be an increasing function of product's quality. In our setup the parameter $\eta_{j,t}$ captures the quality attribute of a product. Thus, the correlation between $\eta_{j,t}$ and the total number of ads aired during the week for each show can serve as a preliminary test of the signaling hypothesis. It turns out that the correlation

36 The table presents the advertising effectiveness for each show aired on Wednesday through Friday and for each network. The effectiveness measure is the predicted increase in the market share of a show (in percentage points) as a result of exposing each individual to an additional ad for the show.

is negative when all the networks are pooled together, and it is negative for 3 networks out of 4 (see Figure 1).[37] This result is robust to various specifications of the model (and thus estimates of $\eta_{j,t}$). Thus, high advertising intensity is not used to signal high quality to rational consumers, ruling out the signaling role of advertising in our data.[38]

5.6. *Robustness check*

We estimate several alternative model specifications to verify the robustness of our results (Tables 8 and 9a–9d). The specific functional forms for these specifications are presented in the notes to Table 8.

Model III. First, we disable the experience signals (set $\varsigma^e = 0$) and re-estimate the model. Since the exposure to experience signals is correlated with the unobservables in preferences, the estimate of ϕ (which determines the effect of experience signals on future choices) is likely to be biased if the unobserved heterogeneity in the model is misspecified. The point estimate of ϕ is -15.86 with a s.e. of 4.46, vs -15.84 ($s.e. = 2.42$) in the baseline. The other parameters are also very similar to the baseline estimates, thus we have no evidence for misspecification of unobserved heterogeneity.

Model IV. Next, we estimate the risk parameter without relying on advertising. We set $\sqrt{\varsigma^a} = 0$, $\upsilon_k^\rho = 0$ (for all k) and re-estimate the model. Since we have variation (across networks and individuals) in the precision of networks' profiles and exposures to experience signals, the risk parameter is still identified. The estimate of ϕ is -15.75 ($s.e. = 2.34$), very similar to the baseline estimate.

Table 9b. Persuasion parameters for Model VI.

	1st ad	2nd ad	3rd ad	4+ ads
Segment 1				
MT	−0.1452 (0.1757)	−0.0077 (0.2269)	−0.6470 (0.3195)	−0.2323 (0.3275)
WF	−0.0772 (0.1217)	−0.2010 (0.1731)	−0.6127 (0.2289)	0.0476 (0.3155)
Segment 2				
MT	0.2362 (0.1674)	0.3113 (0.2271)	0.0564 (0.4102)	0.0004 (0.6559)
WF	−0.0605 (0.1291)	0.3022 (0.1786)	0.3358 (0.2602)	0.2469 (0.3367)
Segment 3				
MT	0.2555 (0.1951)	0.5455 (0.2897)	0.4259 (0.5122)	0.0325 (0.6207)
WF	0.3339 (0.1377)	0.3487 (0.2319)	0.1375 (0.3536)	0.1834 (0.7175)

Note: Standard errors are in parentheses.

37 We focus on Wednesday–Friday shows only, to avoid the measurement problem at the beginning of the week.

38 Moreover, the negative correlations for ABC, CBS and FOX suggest that the networks air more ads to support their lower-quality shows. Thus, if the consumers were using the inference mechanism underlying the signaling effect, they would avoid the heavily-advertised (and hence low-quality) shows.

Table 9c. Persuasion parameters for Model VII.

	ρ_1	ρ_2
Teens	0 (–)	0 (–)
GenX	0.0009 (0.1500)	−0.0750 (0.0497)
Boomer	0.0019 (0.1440)	−0.0317 (0.0462)
Old	0.1294 (0.1561)	−0.0592 (0.0501)
Female	0.0671 (0.0729)	−0.0085 (0.0230)
Income	0.0124 (0.1551)	−0.0516 (0.0470)
Education	−0.1247 (0.1775)	0.0252 (0.0558)
Family	0.0499 (0.0910)	−0.0419 (0.0317)
Urban	−0.0781 (0.0719)	0.0176 (0.0239)
Segment 1		
MT	−0.2470 (0.2536)	0.1006 (0.0697)
WF	−0.1939 (0.2206)	0.1032 (0.0623)
Segment 2		
MT	0.2794 (0.2800)	−0.0155 (0.0889)
WF	0.1087 (0.2393)	0.0995 (0.0666)
Segment 3		
MT	0.2158 (0.3118)	−0.0070 (0.0998)
WF	0.3370 (0.2850)	−0.1425 (0.0972)

Note: Standard errors are in parentheses.

Table 9d. Persuasion parameters for Model VIII.

	Segment 1	Segment 2	Segment 3
MT	−0.2244 (0.1544)	0.1939 (0.1412)	0.3284 (0.1609)
WF	−0.1368 (0.1044)	0.1089 (0.1043)	0.2619 (0.1175)

Note: Standard errors are in parentheses.

Model V. Next, we allow the risk parameter to differ across segments in the baseline specification. The estimates of ϕ_k range from -15.03 to -17.05, and the other estimates are very similar to the baseline, giving a further indication of stability of our estimates.

Finally, we verify that our findings are robust to the functional form of the persuasive effect. To facilitate comparison across different specifications, we assess the magnitudes of the risk-reduction effect and the persuasive effect for each model by measuring the change in market share as a result of exposure to 1, 2 and 3 ads.

First, we estimate a non-parametric specification (Model VI). The risk-reduction effect and the persuasive effect in the non-parametric specification are similar in magnitude to the baseline. Next, we adopt a more flexible parametric specification for the persuasive effect, and allow the coefficients to depend on the demographic variables (Model VII). The demographic coefficients turn out to be insignificant, and the magnitudes of the two effects are close to the baseline. Since an average viewer is not exposed to any ads for over 75% of

Table 10. Advertising effectiveness for Wednesday–Friday shows (percentage points per exposure).

Show	Full model	Risk-neutral model	Show	Full model	Risk-neutral model
ABC			NBC		
Ellen	2.5	3.3	Seaquest 2032	2.1	1.9
The Drew Carey Show	1.2	1.5	Dateline NBC (Wed)	2.5	2.8
Grace Under Fire	2.8	3.5	Friends	5.2	5.7
The Naked Truth	1.3	1.7	The Single Guy	2.2	2.6
Columbo	1.5	2.3	Seinfeld	3.4	3.9
Family Matters	3.2	4.3	Caroline in the City	1.8	2.3
Boy Meets World	1.4	1.8	Unsolved Mysteries	3.0	3.2
Step by Step	1.5	2.1	Dateline NBC (Fri)	1.9	2.1
Hanging with Mr. Cooper	0.9	1.5	Fox		
CBS			Beverly Hills 90210 (Wed)	5.7	5.5
Bless this House	3.4	2.3	Party of Five	1.7	0.4
Dave's World	2.1	2.2	Living Single	3.5	2.4
Central Park West	1.7	0.7	The Crew	1.6	1.4
Murder, She Wrote	3.0	3.3	New York Undercover	2.0	1.8
New York News	0.9	0.7	Strange Luck	4.1	1.5
Here Comes the Bride	3.4	1.9	X-Files	4.7	5.0
Ice Wars	2.1	1.4			
Average advertising effectiveness					
ABC	1.8	2.5	NBC	2.8	3.1
CBS	2.4	1.8	FOX	3.3	2.6

the shows, we also estimate a specification which only distinguishes between non-exposure and exposure to any positive number of ads in the persuasive effect (Model VIII). Again, the results are similar to the baseline. Finally, we get similar results on the importance of the risk-reduction effect vis-a-vis the direct effect of ads in Models III and V. Recall that these models use the baseline specification of the persuasive effect but modify other elements of the model.[39,40]

39 We've also tried an alternative semi-structural specification of risk preferences $\tilde{V}(\mu^p_{i,j,t}, \sigma^2_{i,j,t}) = \mu^p_{i,j,t} + \phi\sigma^2_{i,j,t}$. That is, we replaced the standard deviation of the posterior with the variance. The likelihood is substantially worse than in the baseline, but we still find strong risk-aversion and negligible persuasive effect.

40 We have also tried to extend the information structure of the model in the following way. Instead of observing the true value of the show-individual random effect $\nu_{i,j,t}$, the individuals try to infer it from the prior distribution, a miscellaneous signal, experience signals and advertising signals (all these additional signals are orthogonal to the original signals from the baseline specification). Otherwise the model is similar to the baseline. Thus, the model includes learning of both product characteristics observable to the researcher (as in the baseline) and show-individual unobservables. We cannot reject the null of perfect information about $\nu_{i,j,t}$ (i.e. the baseline specification) at the 5% level, and the estimates of the direct effect and the risk-reduction effect are close to the baseline estimates.

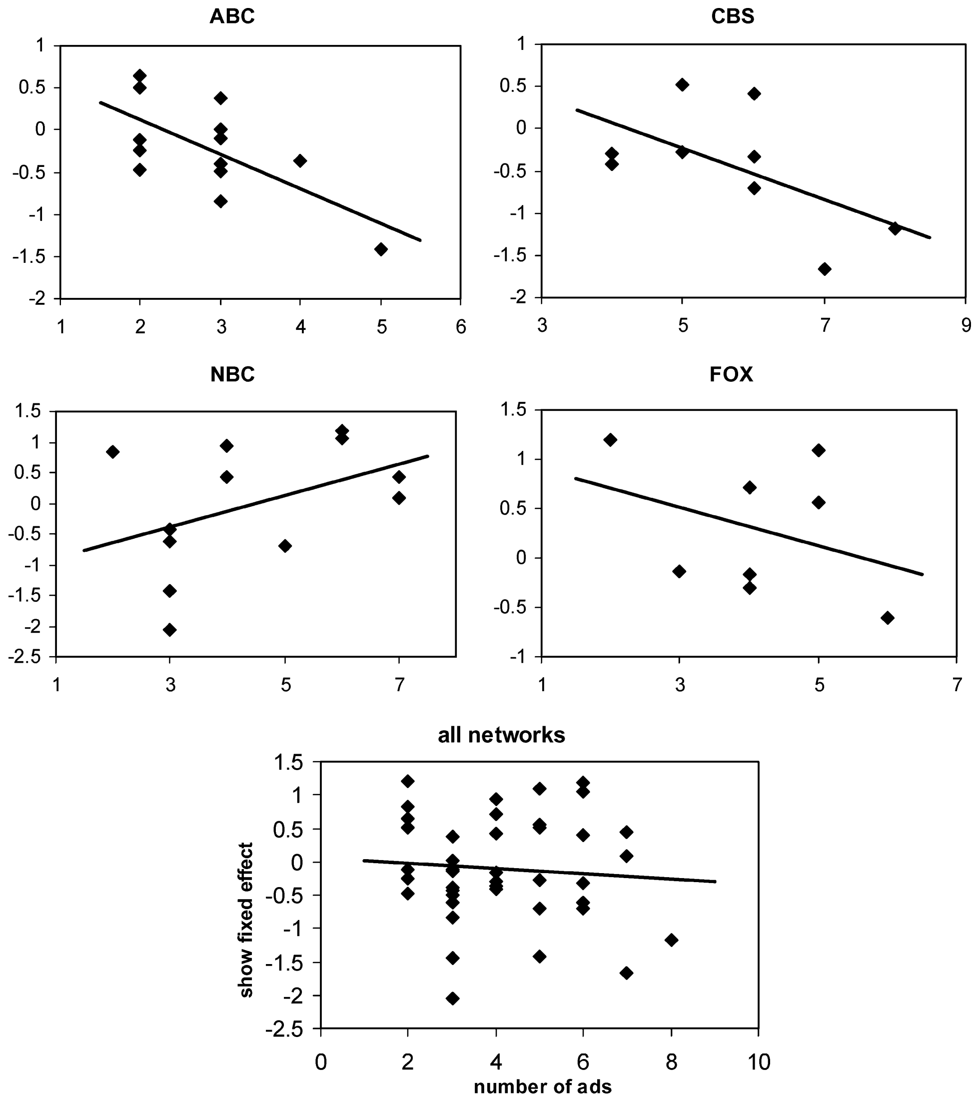

Figure 1. The total number of ads aired for a show and the show fixed effect, $\hat{\eta}_{j,t}$.

6. Conclusion and managerial implications

This study shows theoretically and empirically that exposure to advertising increases consumers' tendency to purchase the promoted product *because* the informative content of advertising resolves some of the uncertainty that the risk-averse consumers face and thus reduces the risk associated with the product.

Our model implies that advertising effectiveness depends on (a) the risk preference parameter, (b) the precision of the advertising message, (c) the familiarity of the consumer with the product, (d) the consumer's involvement level with the product, and (e) the diversity of products offered by multiproduct firms.

These findings have significant managerial implications. Advertising strategies are determined based on ad effectiveness. Specifically, two of the most important decisions of ad agencies depend on ad effectiveness—ads spending and audience targeting. Our model suggests that ads spending should be higher (a) for new and relatively unknown products, (b) for high-involvement products, (c) when ads can be quite precise, and (d) when the firm offers a diverse product-line. It can also assist targeting strategies. Ads should target consumers (a) who are more sensitive to risk—for example, women versus men (see Palsson, 1996; Jianakoplos and Bernasek, 1998), (b) who are more sensitive to products' attributes (and thus more involved), and (c) those who are not familiar with the promoted product.

The empirical result, that advertising has a significant risk-reduction role but negligible direct effect on the utility, might be important for models of advertising competition. Currently, most such models assume that advertising has a direct (persuasive) effect on consumers' utility.

However, the empirical result, that advertising has a significant risk-reduction role but negligible direct effect on the utility, might be specific to our data set and the empirical example (where the products are television shows and ads are promos for these shows). The generalizability of this result should be examined by estimating the risk-reduction model with additional data sets. Furthermore, the model can be examined empirically also by directly testing its implication. Previous studies have already established the dependence of advertising effectiveness on consumers' familiarity with products. Furthermore, experimental studies (e.g. Miniard et al., 1991) found that the impact of ads containing product-relevant information is stronger for the high-involvement consumers. The predictions of the risk-reduction model are consistent with these findings, and offer a new explanation for them.

The welfare implications of the model are quite straightforward. The positive correlation between exposure to advertising and consumption was established long ago. This empirical regularity disturbed scholars (Galbraith, 1958, 1967; Packard, 1957, 1969; Telser, 1964; Stiglitz, 1989) for a while. The main reason for the concern was the potential persuasive power of advertising. It was interpreted as if consumers can be manipulated by ads. Advertising was accused of creating wants, distorting tastes and persuading consumers to buy products that they do not need. Our findings calm these concerns a bit, since (at least in our empirical example) advertising does not have a direct effect on consumers' utility. Furthermore, instead of having these negative aspects, advertising in our model informs consumers about product attributes and as a result reduces the undesirable tension that is associated with uncertainty. Furthermore, its role according to the risk-reduction model might be limited. For example, unlike the standard full-information approach, our model predicts that advertising can hardly increase the sales of a well-known product.

Finally, our findings about the significant risk parameter might be relevant for choice models in general (not only for advertising models). Imperfect information is included in recent empirical studies more often than in the past. However, many of these studies (obviously not all of them) still assume that consumers are risk-neutral. This assumption is probably ungrounded. Furthermore, this study demonstrates that by accounting for consumers' risk aversion, one can avoid inconsistent estimates and misleading implications.

References

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

Ackerberg, D. (2003). "Advertising, Learning and Consumer Choice in Experience Good Markets: A Structural Empirical Examination." International Economic Review 44(3), 1007–1039.

Anand, B. and R. Shachar. (2004). "Brands as Beacons: A New Source of Loyalty to Multiproduct Firms." Journal of Marketing Research 41(2), 135–150.

Anand, B. and R. Shachar. (2001). "Advertising, the Matchmaker." The Israeli Institute of Business Research Working paper No. 19/2001, December 2001. (http://www.tau.ac.il/˜rroonn/Papers/match.pdf).

Batra, R., J.G. Myers, and D.A. Aaker. (1996). Advertising Management, 5th edition. New Jersey: Prentice-Hall International.

Becker, G. and K. Murphy. (1993). "A Simple Theory of Advertising as a Good or Bad." Quarterly Journal of Economics 108, 941–963.

Braithwaite, D. (1928). "The Economic Effects of Advertisement." Economic Journal 38, March, 16–37.

Darmon, R.Y. (1976). "Determinants of TV Viewing." Journal of Advertising Research 16, 17–24.

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

Erdem, T. and B. Sun. (2002). "An Empirical Investigation of Spillover Effects of Marketing Mix Strategy in Umbrella Branding." Journal of Marketing Research 39(4), 408–420.

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

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

Galbraith, J. (1958). The Affluent Society. Houghton Mifflin.

Galbraith, J. (1967). The New Industrial State. Houghton Mifflin.

Grossman, G. and C. Shapiro. (1984). "Informative Advertising with Differentiated Products." Review of Economic Studies 51, 63–81.

Goettler, R. and R. Shachar. (2001). "Spatial Competition in the Network Television Industry." RAND Journal of Economics 32(4), 624–656.

Hajivassiliou, V. (1997). "Some Practical Issues in Maximum Simulated Likelihood." London School of Economics Discussion Paper Series EM/97/340.

Heckman, J. (1981). "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time—Discrete Data Stochastic Process." In C. Manski and D. McFadden (eds.), Structural Analysis of Discrete Data with Econometric Applications, Cambridge: MIT Press.

Horen, J.H. (1980). "Scheduling of Network Television Programs." Management Science 26, 354–370.

Jacoby, Jacob and Wayne Hoyer. (1982). "Viewer Miscomprehension of Televised Communication: Selected Findings." Journal of Marketing 46(4), 12–26.

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

Jianakoplos, N. and A. Bernasek. (1998). "Are Women More Risk Averse?" Economic Inquiry 36(4), 620–630.

Kirmani, Amna, and Peter Wright. (1989). "Money Talks: Perceived Advertising Expense and Expected Product Quality." Journal of Consumer Research 16 (December), 344–353

McFadden, D. (1973). "Conditional Logit Analysis of Qualitative Choice Behavior." In P. Zagrembka (ed.), Frontiers in Econometrics, New York: Academic Press.

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

Miniard, P., S. Bhatla, K. Lord, P. Dickson, and H. Unnava. (1991). "Picture-Based Persuasion Processes and the Moderating Role of Involvement." Journal of Consumer Research 18(1), 92–107.

Moshkin, Nickolay and Ron Shachar. (2002). "The Asymmetric Information Model of State Dependence." Marketing Science 21(4), 435–454.

Nelson, Philip. (1974). "Advertising as Information." Journal of Political Economy 82(4), 729–754.

Nevo, A. (2001). "Measuring Market Power in the Ready-to-Eat Cereal Industry." Econometrica 69(2), 307–342.

Packard, V. (1957). Hidden Persuaders. New York: D. McKay Co.

Packard, V. (1969). The Waste Makers. New York: D. McKay Co.

Palsson, Anne-Marie. (1996). "Does the Degree of Relative Risk Aversion Vary with Household Characteristics." Journal of Economic Psychology 17(6), 771–787.

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.

Rust, R.T. and M.I. Alpert. (1984). "An Audience Flow Model of Television Viewing Choice." Marketing Science 3, 113–124.

Shachar, R. and J. Emerson. (2000). "Cast Demographics, Unobserved Segments and Heterogenous Switching Costs in a TV Viewing Choice Model." Journal of Marketing Research 37, 173–186.

Shachar, R. and B. Anand. (1998). "The Effectiveness and Targeting of Television Advertising." Journal of Economics and Management Strategy 7(3), 363–396.

Stern, S. and M. Trajtenberg. (1998). "Empirical Implications of Physician Authority in Pharmaceutical Decision-making," NBER Working Paper No. 6851.

Stiglitz, J. (1989). "Reflections on the State of Economics: 1988." Economic Record, March 1989, 66–72.

Tellis, Gerard J., Rajesh Chandy, and Pattana Thaivanich. (2000). "Decomposing the Effects of Direct Advertising: Which Brand Works, When, Where, and How Long?" Journal of Marketing Research 37 (February), 32–46.

Telser, L. (1964). "Advertising and Competition." Journal of Political Economy 72(6), 537–562.

Tirole, Jean. (1988). The Theory of Industrial Organization. Cambridge, MA: The MIT Press.