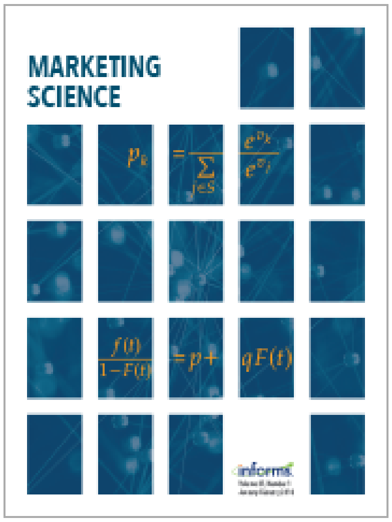


Marketing Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

The Effect of Media Advertising on Brand Consideration and Choice

Nobuhiko Terui, Masataka Ban, Greg M. Allenby,

MARKETING SCIENCE
CELEBRATING 30 YEARS
Vol. 30, No. 1, January–February 2011, pp. 74–91
ISSN 0732-2399 | EISSN 1526-548X | 11 | 3001 | 0074
DOI 10.1287/mksc.1100.0590

The Effect of Media Advertising on Brand Consideration and Choice

Nobuhiko Terui
Graduate School of Economics and Management, Tohoku University, Sendai 980-8576, Japan, terui@econ.tohoku.ac.jp

Masataka Ban
Faculty of Business Administration, Mejiro University, Tokyo 161-8539, Japan, ban@mejiro.ac.jp

Greg M. Allenby
Fisher College of Business, Ohio State University, Columbus, Ohio 43210, allenby.1@osu.edu

The nature of the effect of media advertising on brand choice is investigated in two product categories in analyses that combine household scanner panel data with media exposure information. Alternative model specifications are tested in which advertising is assumed to directly affect brand utility, model error variance, and brand consideration. We find strong support for advertising effects on choice through an indirect route of consideration set formation that does not directly affect brand utility. Implications for media buying and advertising effects are explored.

Key words: Bayesian analysis; threshold effects; determinants of utility
History: Received: October 7, 2009; accepted: May 25, 2010; Eric Bradlow served as the editor-in-chief and Alan Montgomery served as associate editor for this article. Published online in *Articles in Advance* September 24, 2010.

1. Introduction

A variety of mechanisms have been proposed in the marketing literature to study the effect of advertising on brand choice. Advertising can inform consumers about product benefits and can influence consumer brand evaluations and subsequent purchase decisions (Mehta et al. 2008). Advertising effects are measured with models where consumers are assumed to learn about the true quality of a product over time (Erdem and Keane 1996) and with models where advertising is assumed to affect model coefficients, such as inducing consumers to be more price sensitive (Jedidi et al. 1999). The most predominant effect reported in the literature is one that has a direct effect on sales and profitability (Mizik and Jacobson 2008, Ansari and Mela 2003, Feinberg 2001, Erickson 1997, Lodish et al. 1995). Early studies of advertising effects relied on aggregate data and time-series models to estimate carryover effects (Clarke 1976), and more recent studies have investigated advertising effects using scanner panel data (Pedrick and Zufryden 1991, Horsky et al. 2006). Researchers have documented the presence of interactive effects of advertising on price sensitivity and other variables (Mela et al. 1997, Kanetkar et al. 1992), and more recently, they have proposed threshold models to capture the effect of advertising in aggregate data, where consumer responses change when advertising levels are sufficiently high (Dubé et al. 2005, Vakratsas et al. 2004). An implicit theme in these analyses is that advertising has a direct and possibly lagged effect on consumer utility, either by informing consumers of brand characteristics or by making them more sensitive to existing characteristics.

In this paper, we provide empirical evidence that challenges the assumption that media advertising has a direct effect on consumer utility for established, or mature, brands. We present results from an analysis of two scanner panel data sets, laundry detergent and instant coffee, where results indicate that household exposure to media advertising affects the construction of consideration sets, not the marginal utility of offerings. Once the effects of advertising on consideration set formation are accounted for, the added effects on utility are found to be negligible. Thus, our results support the use of brand consideration as a dependent variable in studying advertising effects (Naik et al. 1998, Villas-Boas 1993) and support a growing literature on the presence of multistaged decision processes (Chandukala et al. 2009, Chessa and Murre 2007) originally articulated in early research on advertising effects (e.g., Howard and Sheth 1969, Lavidge and Steiner 1961, Strong 1925). Analysis that attempts to establish a direct association with sales is found to underestimate advertising effectiveness.

Our results are consistent with experimental evidence demonstrating that the information used to screen choices for consideration becomes less important than other information acquired later in the search process (Chakravarti et al. 2006). However, our results are not in agreement with past empirical research focusing on the role of advertising on marketplace choices. Andrews and Srinivasan (1995), for example, find that media advertising affects both brand consideration and choice, and as mentioned previously, past research has focused on finding direct effects. We believe there are two reasons for the difference between our results and past empirical studies. First, our analysis allows for heterogeneous consumer response to advertising and consideration set formation. Most prior studies have not allowed for the presence of heterogeneous consideration sets with marketplace data. Second, our model imposes a hard constraint on brand inclusion in the consideration set, in contrast to models that introduce a separate probability for consideration set inclusion (e.g., Bronnenberg and Vanhonecker 1996). We believe that the presence of a hard constraint is more consistent with the concept of a consideration set where alternatives are either considered or not, leading to a more sharply defined model likelihood. The value of hard constraints has previously been demonstrated by Gilbride and Allenby (2004). We find that a hard constraint model provides a significantly better fit to the data by clearly distinguishing consideration from choice in the model likelihood, allowing us to associate the impact of advertising on brand consideration.

Consumer exposure to media advertising is assumed to contribute to an accumulation of advertising stock that exponentially decays over time. Brands enter the household consideration set if advertising stock is greater than a threshold value. The decay rate and threshold value are parameters that are estimated heterogeneously. Our consideration set model is shown to fit the data better than models that directly associate advertising exposures to a brand's marginal utility and indirectly through the variance of random utility error. We also provide evidence that our model outperforms earlier models of brand consideration and consumer learning of quality by advertising signals (Erdem and Kean 1996). An implication of our model structure is that just one (periodic) exposure is sufficient for consideration set inclusion by raising the advertising stock over the threshold level. We also show that the presence of consideration sets leads to asymmetric advertising effects, where decreased advertising can lead to large losses in sales as brands are excluded from the consideration set.

The remainder of this paper is organized as follows. In §2 we develop our model, which builds on the consideration set model developed by Gilbride and Allenby (2004), by introducing heterogeneous latent advertising stock variables that temporally affect household consideration sets. The data used in the analysis are described in §3, and §4 presents empirical results. Implications are discussed in §5, and concluding remarks are offered in §6.

2. Model Development

We begin by assuming that there are N choice alternatives, and consumer h has advertising stock AS_{jht} for each alternative $(j = 1, \ldots, N)$ throughout the observational period. Alternative j enters the consideration of consumer h at time t when advertising stock AS_{jht} is larger than a threshold value r_h that is common across choice alternatives and constant through time. Let $C_{ht}^{[AS_{jht} \geq r_h]}$ indicate the set of choice indices having advertising stock greater than r_h:

$$\text{if } AS_{jht} \geq r_h, \text{ then } j \in C_{ht}^{[AS_{jht} \geq r_h]}. \tag{1}$$

We define $C_{ht}^{[AS_{jht} \geq r_h]}$ as a consideration set of consumer h at time t. We note that the number of elements of $C_{ht}^{[AS_{jht} \geq r_h]}$ changes over time with changes in the stock variable.

Next, we assume that consumer h has a utility function for the alternative when it enters the consumer's consideration set:

$$u_{jht} = x'_{jht}\beta_h + \varepsilon_{jht}, \quad \varepsilon_{jht} \sim \mathrm{N}(0, \sigma_j^2 = 1) \quad \text{for } j \in C_{ht}^{[AS_{jht} \geq r_h]}. \tag{2}$$

We note that the vectors of errors in Equation (2) could be specified to have a general covariance matrix subject to standard identifying restrictions. We instead choose to work with a restricted covariance matrix in our model (i.e., $\sigma_j^2 = 1.0$), and we allow for departures in alternative, competitive models that are described in the following. Based on (1) and (2), the choice probability at time t for the alternatives in the consideration set as a function of advertising stock is

$$\Pr(j)_{ht} = \Pr\left\{u_{jht} = \max\left\{u_{kht}\colon k \in C_{ht}^{[AS_{jht} \geq r_h]}\right\}\right\}. \tag{3}$$

Our model is logically constrained in that observed choices are assumed to come from the set of considered brands. This implies that if a respondent is not exposed to a brand's advertisements but is observed to purchase the brand, then the threshold value r_h must be near 0. Whenever the advertising stock variable is less than the threshold value, the choice probability is equal to 0.

The final component of our model deals with the formation and depletion of advertising stock and brand loyalty stock. Following Bass and Clark (1972), Clarke (1976), Guadagni and Little (1983), and others, the stock variables are defined as an exponentially weighted average of past exposures:

Advertising Stock (AS):

$$AS_{jht} = \sum_{g=0}^{\infty} a_{jht-g}\rho_h^g \quad (0 \le \rho_h < 1), \tag{4}$$

where a_{jht-g} is a variable indicating that respondent h was exposed to media advertising for brand j at time $t-g$, and ρ_h is a parameter reflecting advertising diminishing effect for $0 \le \rho_h < 1$. Thus, the effect of advertising is assumed to occur instantly and decay exponentially. The presence of exponential decay in models of advertising is consistent with previous research (Little 1979, Lodish et al. 1995). The decay parameter ρ_h is specified as being household specific.

Our model is consistent with other aggregate models that have been used to study advertising effects—e.g., Lambin (1976), Simon and Arndt (1980), Bemmaor (1984), Bronnenberg (1998), Steiner (1987), Vakratsas and Ambler (1999), Hanssens et al. (2001), and Vakratsas et al. (2004). In these papers, the threshold r_h is called the "effective advertising stock." However, in contrast to these models, we assume advertising works through the formation of an individual's consideration set, not thresholds designed to enhance its direct effect on a brand's utility. An alternative formulation is based on a share of voice assumption that normalizes the data so that advertising is expressed as a proportion. We do not pursue this option because the normalization neglects the level of advertising intensity. Advertising has also been studied using models of learning (e.g., Erdem and Keane 1996, Erdem et al. 2008, Mehta et al. 2008), where nonlinearity arises from Bayesian updating of quality signals. In these learning models, the effects of advertising are reflected in changes to the marginal utility of the brand, not the formation of a consideration set.

We investigate a number of alternative specifications for consideration set formation and describe them in detail in the next section. These include models where consideration sets are formed on the basis of past purchases and merchandising (i.e., display) variables instead of advertising exposure. We also investigate a class of models in which consideration sets are formed stochastically. A distinctive aspect of our model is that it imposes a hard constraint on consideration set formation—e.g., brands with advertising stock incrementally below the threshold are immediately excluded from consideration. We examine the reasonableness of this assumption by fitting a stochastic consideration set model and models in which the scale of the random utility error is related to various stock (Allenby and Ginter 1995). We also examine the performance of various models where merchandising and advertising variables are directly entered into the model specification to modify the marginal utility of the offerings, including the Bayesian updating model of Erdem and Keane (1996). Finally, our model of brand consideration does not preclude the further narrowing of the consideration set by, for example, price search behavior (see Mehta et al. 2003). Our model can be viewed as providing a starting point for structural models of consideration for price.

Heterogeneity is introduced into the model parameters through a random-effects specification. The resulting hierarchical model is estimated using Markov chain Monte Carlo methods that are described in the appendix.

3. Data and Alternative Models

Two scanner panel data sets of laundry detergent and instant coffee purchases are provided by Video Research Inc., Japan. The detergent data comprise six leading brands during a 100-week period, beginning with the 14th week of 1990 and ending with the 9th week of 1992. From 98 households, we obtained 9,900 records for our analysis. The coffee data comprise four leading brands during October 1, 1993 through March 31, 1994 (183 days), with 237 households present in the panel. The first 70 weeks for detergent data and 152 days for coffee data are used for model estimation, and the remaining periods are reserved for holdout predictions. In addition to purchase data, data are available that record household exposure to television advertising for the brands. The advertising data are collected automatically using an in-home recording device connected to household television sets.

Three variables are used as covariates in the analysis: advertising $\{a_{jht}\}$, price $\{p_{jht}\}$, and in-store displays $\{d_{jht}\}$, where j denotes the brand, h denotes the household, and t denotes the time of purchase. We also investigate consideration sets based on two other stock variables; the first is as follows:

Brand Loyalty (BL):

$$BL_{jht} = \sum_{g=1}^{\infty} y_{jht-g}\tau_h^g \quad 0 \le \tau_h < 1, \tag{5}$$

where y_{jht-g} denotes the purchase variable for brand j, household h, and time $(t-g)$ that takes on values of 1 and 0. The brand loyalty variable construction follows Guadagni and Little (1983) and others studying state dependence carryover effects (Erdem 1996).

The second variable, display stock variable, is constructed in a similar manner:

Display Stock (DS):

$$DS_{jht} = \sum_{g=0}^{\infty} d_{jht-g}\phi_h^g \quad 0 \le \phi_h < 1. \tag{6}$$

Both variables are assumed to exponentially decline through time, each with their own threshold parameter (λ_h and κ_h, respectively).

Table 1 displays summary statistics of prices, display frequencies, and advertising counts for each of the brands. Laundry detergent brand 6 and coffee brand 1 have the largest choice share and number of advertising exposures, as well as the highest rates of price discount and display activity. More interesting, however, are the low-share brands in both data sets—brands 4 and 5 in the detergent data and brands 2 and 3 in the coffee data. These brands differ in their advertising and display activity, with detergent brand 5 and coffee brand 2 engaging in much greater use of advertising and fewer displays than detergent brand 4 and coffee brand 3. These smaller-share brands offer the chance to distinguish models of consideration based on advertising, brand loyalty, and display stock. Brands in the coffee category advertise at a much higher rate than brands in the detergent category, and the coffee expected interpurchase time is about 20% shorter than that for detergent.

Figure 1 displays the time series of advertising exposures for each of brands in our analysis. Panel a shows the plots for the detergent brands, and panel b shows the plots for coffee. Plotted is the total number of exposures to all panel members within each week of the data set. The plots indicate that advertising levels were not growing or declining over time. In addition, we conducted an aggregate vector autoregressive analysis of these data along with display and price levels to investigate the potential presence of strategic timing of the variables. Across all brands and variables, the only consistent relationship detected was a one-period lagged effect indicating that increases in one period are most likely followed by a decline in the next period for each of the individual variables. This can be seen visually in Figure 1 for sales and advertising exposure, where an increase in one period is often followed by a decrease in the next. Cross-correlation analysis failed to detect any lead-lag relationship among the aggregated variables, indicating that the timing of media insertions is unrelated to aggregate sales levels or the timing of merchandising variables. These results cast doubt on advertising levels being strategically timed to coincide with favorable demand conditions.

A second reason for questioning the presence of endogenous advertising levels is due to the scheduling practices of advertisers for these data. There is a six-month lag between the time a contract is signed and an advertisement is aired, implying that firms would need to predict the presence of demand shocks far in advance for temporal endogeneity to be present. Endogeneity, however, may be present in the cross-sectional allocation of advertising budgets, but this will not lead to inconsistent parameter estimates in our analysis if the allocation to the cross-sectional units is constant—i.e., an endogenous effect can only be an effect if it varies (Rubin 1976, Liu et al. 2007). Because our data pertain to households within a specific region, and advertising is distributed en masse, we believe that the potential impact of advertising endogeneity is minimized and can be ignored.

Alternative models used to investigate advertising effects are categorized as having either a direct or indirect advertising effect on utility. We investigate a wide variety of models for comparison and to ascertain the robustness of our results. The direct effect models are grouped into linear utility function models with and without advertising stock as well as into nonlinear utility function models with squared and cubed terms of advertising stock and brand loyalty to capture the nonlinear response of advertising discussed in the literature. These higher-order polynomials of stock variables provide a flexible approach

Table 1 Description of the Data

Laundry detergent	Choice share	Total advertising exposures	Average display frequency	Average price (discount rate)	Average interpurchase time (weeks)
Brand 1	0.241	1,183	0.278	0.712	17.291
Brand 2	0.133	976	0.065	0.763	11.515
Brand 3	0.122	416	0.042	0.823	16.543
Brand 4	0.069	183	0.226	0.919	11.570
Brand 5	0.053	599	0.139	0.919	12.583
Brand 6	0.382	1,632	0.364	0.662	15.220
Instant coffee	**Choice share**	**Total advertising exposures**	**Average display frequency**	**Average price (per 100 g)**	**Average interpurchase time (days)**
Brand 1	0.523	34,226	0.058	450.018	43.616
Brand 2	0.132	29,749	0.007	406.484	54.510
Brand 3	0.111	9,189	0.016	375.465	46.975
Brand 4	0.234	6,847	0.020	470.392	45.704

Notes. For laundry detergent, sample size: 98 households, 70 weeks, 730 purchase occasions. For instant coffee, sample size: 237 households, 151 days, 551 purchase occasions.

Figure 1 Ad Exposure and Sales Time-Series Plots (Aggregated)

(a) Detergent

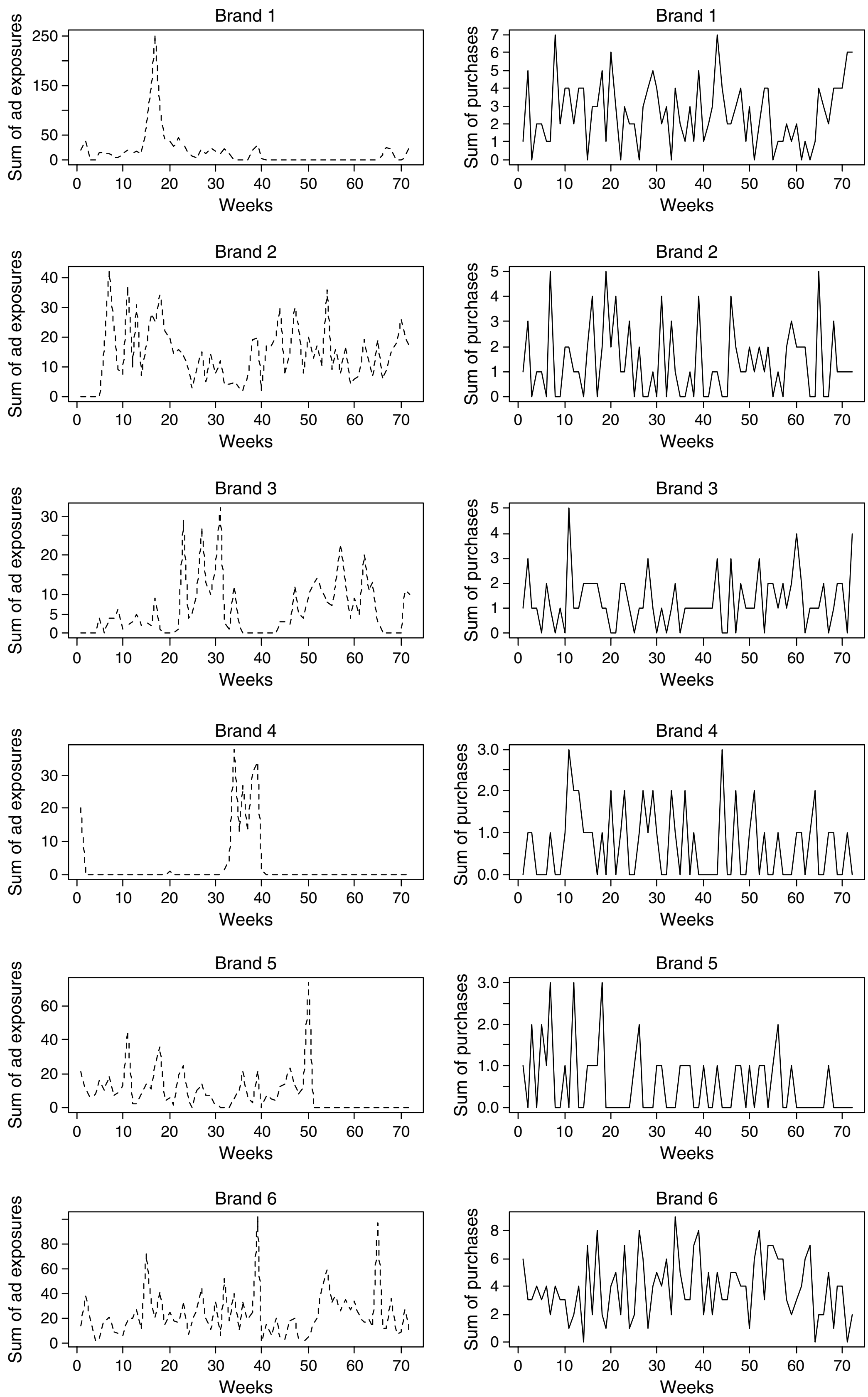
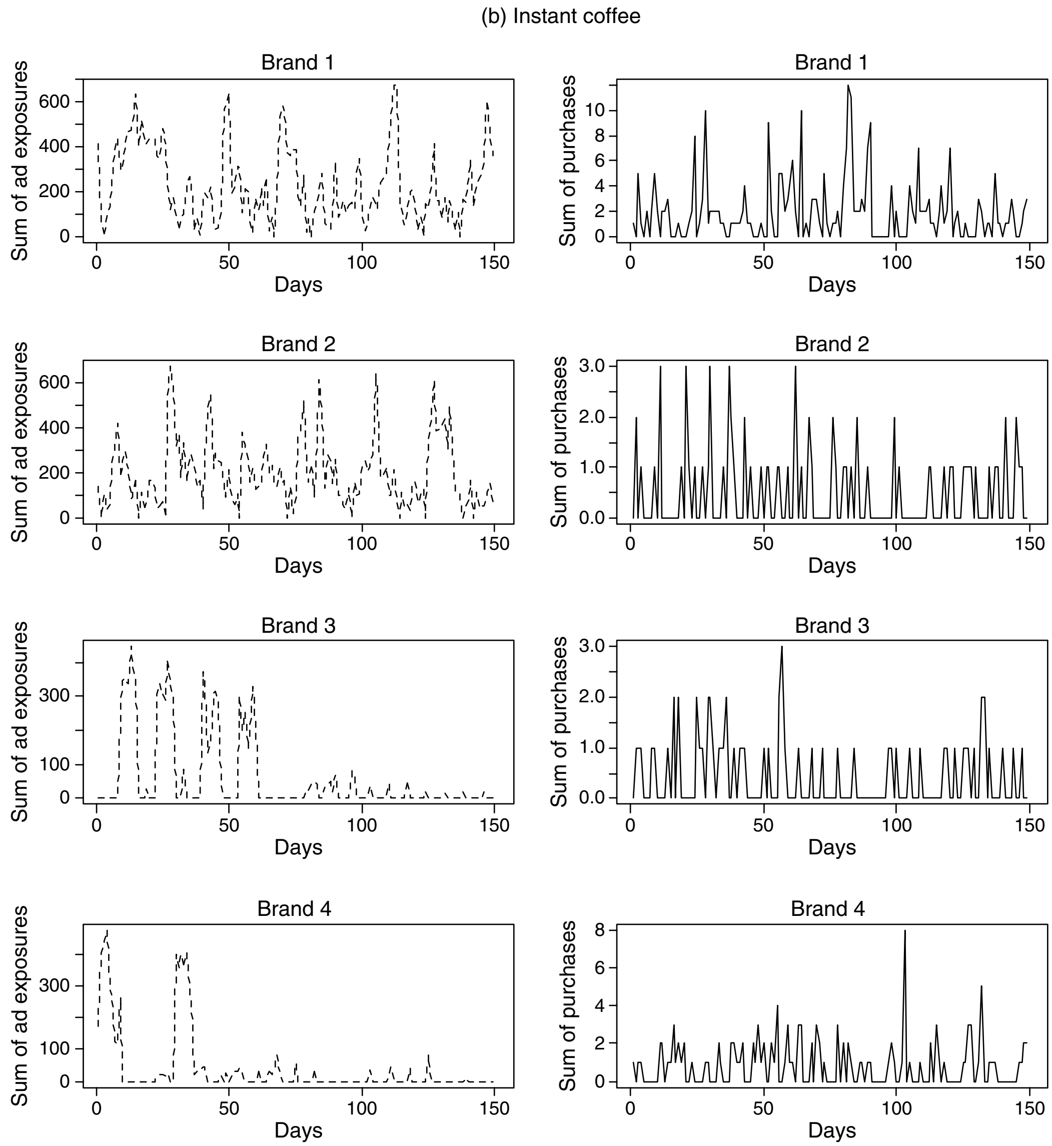


Figure 1 (Cont'd.)



to understanding time-varying effects of advertising assumed to affect household-marginal utility of the brand. Thus, in our analysis, we attempt to approximate time-varying effects rather than to employ specific functional forms as with, for example, models of learning (e.g., Mehta et al. 2008). In the specifications below, we denote $M_{jht} = \{\text{brand intercept}_j,\ p_{jht},\ d_{jht}\}$ as brand intercept, price, and display variables entered into the models in the traditional manner.

Direct Effect Models

Linear

N0: $u = f(M);\quad u_{jht} = M'_{jht}\beta_h + \varepsilon_{jht}.$

N1: $u = f(M, AS);\quad u_{jht} = M'_{jht}\beta_h^{(1)} + AS_{jht}\beta_h^{(2)} + \varepsilon_{jht}.$

N2: $u = f(M, AS, BL);$

$$u_{jht} = M'_{jht}\beta_h^{(1)} + AS_{jht}\beta_h^{(2)} + {}_{jht}\beta_h^{(3)} + \varepsilon_{jht}.$$

N3: $u = f(M, AS, BL, BL^2);$

$$DS\,u_{jht} = DSM'_{jht}\beta_h^{(1)} + AS_{jht}\beta_h^{(2)} + BL_{jht}\beta_h^{(3)} + BL_{jht}^2\beta_h^{(4)} + \varepsilon_{jht}.$$

N4: $u = f(M, AS, BL, BL^2, BL^3);$

$$u_{jht} = M'_{jht}\beta_h^{(1)} + AS_{jht}\beta_h^{(2)} + {}_{jht}\beta_h^{(3)}AS + BL_{jht}^2\beta_h^{(4)}.$$
$$+ BL_{jht}^3\beta_h^{(5)} + \varepsilon_{jht}.$$

Nonlinear

N5: $u = f(M, AS, AS \times \text{Price});$

$$u_{jht} = M'_{jht}\beta_h^{(1)} + AS_{jht}\beta_h^{(2)} + AS_{jht} \times \text{Price}_{jht}\beta_h^{(3)} + \varepsilon_{jht}.$$

N6: $u = f(M, AS, AS^2);$

$$u_{jht} = M'_{jht}\beta_h^{(1)} + AS_{jht}\beta_h^{(2)} + AS_{jht}^2\beta_h^{(3)} + \varepsilon_{jht}.$$

N7: $u = f(M, AS, AS^2, AS^3);$

$$u_{jht} = M'_{jht}\beta_h^{(1)} + AS_{jht}\beta_h^{(2)} + AS_{jht}^2\beta_h^{(3)} + AS_{jht}^3\beta_h^{(4)} + \varepsilon_{jht}.$$

We also consider indirect effect models in which advertising and other stock variables do not directly affect utility but instead work to form a considered set of brands for purchase. We consider a variety of models of consideration set formation based on advertising stock, brand loyalty stock, and display exposure stock, as well as various combinations of these variables.

Indirect Effect Models

Consideration Sets via Screening Rules

S0: $u = f(M, AS) \mid scr(AS)$.

S1: $u = f(M) \mid scr(AS)$.

S2: $u = f(M) \mid scr(BL)$.

S3: $u = f(M) \mid scr(DS)$.

S4: $u = f(M) \mid (scr(AS) \text{ or } scr(BL))$.

S5: $u = f(M) \mid (scr(AS) \text{ or } scr(DS))$.

S6: $u = f(M) \mid (scr(AS) \text{ or } scr(BL) \text{ or } scr(DS))$.

S7: $u = f(M) \mid (scr(AS) \text{ and } scr(BL))$.

S8: $u = f(M) \mid (scr(AS) \text{ and } scr(DS))$.

S9: $u = f(M) \mid (scr(AS) \text{ and } scr(BL)) \text{ and } scr(DS))$.

As noted earlier, "M" comprises intercept, price, and display terms, and in some models, the display variable is also used to form the screening rule. Thus, display is always present in the model, with an immediate effect on marginal utility, and is also sometimes present with a lagged effect on consideration set formation.

Finally, we consider the effect of two adjustments to our model. The first is to allow for heterogeneous, brand-specific thresholds r_{hk} for consideration set inclusion. An advantage of this specification is that it effectively reduces the sensitivity of the results to the initial value of the stock variables in Equations (4)–(6). For example, high values of r_{hk} can be used to reflect high initial levels of advertising stock that dissipate over time. For brands that infrequently advertise but are still purchased, the brand-specific threshold can be made arbitrarily small so that purchased brands have a nonzero probability of purchase. In the extreme case where brands do not advertise at all, the brand-specific threshold will be zero, and our choice model will revert to a standard random utility model. The disadvantage is that it complicates the interpretation of the parameters as r_{hk} is no longer just associated with advertising effects only. The second adjustment is to allow for a brand loyalty variable as part of the set of covariates in M; i.e., $M_{jht} = \{\text{brand intercept}_j, p_{jht}, d_{jht}, BL_{jht}\}$. This specification allows for the presence of state dependence.

The goal of our analysis is to assess the role of media advertising on brand consideration and choice. To this end, we are particularly interested in the relationship between models N0, N1, S0, and S1. Model S0 allows advertising stock to directly affect brand utility, in addition to potentially affecting the formation of a consideration set. Model S1 does not allow for a direct effect. A comparison of the fit statistics associated with models S0 and S1 allows assessment of the incremental effect of advertising stock on utility given its role in consideration set formation. Further comparison to models N0 and N1 allows for assessment of the effects of advertising stock on consideration set formation. If the fit of model N1 is greater than N0, and also about the same as S0 and S1 (i.e., N1 = S0 > N0 = S1), then we can conclude that advertising affects brand utility but not consideration set formation. However, if we find that the fit of S1 is about the same as S0 but much better than N1 (i.e., S0 = S1 > N1), then advertising affects brand consideration but not utility. Finally, if the fit of model S0 is better than S1, and S1 is better than N1 (i.e., S0 > S1 > N1 > N0), then we would conclude that advertising affects both brand consideration and utility.

The other screening rule models (S2–S9) examine whether advertising stock is the best variable for consideration set formation and whether these sets are formed by different combinations of stock variables. We examine both conjunctive ("and") and disjunctive ("or") combinations that allow exploration of possible interactive effects.

A potential criticism of the consideration set model described in (3) is that it imposes sharp boundaries on brand inclusion. For example, an advertising stock value AS_{jht} slightly below the threshold value r_h results in strict exclusion from the consideration set and an assigned choice probability of 0. An alternative approach is to impose soft boundaries on consideration set inclusion. We explore two approaches for accomplishing this.

The first approach comes from Bronnenberg and Vanhonaker (1996) and others who posit a separate probability for consideration set inclusion. We introduce a latent-state variable z_{jht}, taking a value of 1 if brand j is included in the consideration set (i.e., $AS_{jht}(\rho_h) \geq r_h$), and 0 otherwise. The likelihood for z_{jht} is

$$\begin{aligned}\Pr\{z_{jht}=1\} &= p_{jht} = \frac{1}{1+\exp\{-\delta_h(AS_{jht}(\rho_h)-r_h)\}},\\ \Pr\{z_{jht}=0\} &= 1-p_{jht} = \frac{\exp\{-\delta_h(AS_{jht}(\rho_h)-r_h)\}}{1+\exp\{-\delta_h(AS_{jht}(\rho_h)-r_h)\}},\end{aligned} \tag{7}$$

where δ_h is the scale adjustment parameter. We show in the appendix that this model results in an adjustment to the variance of the probit model. We denote this model as a stochastic screening (SS) rule model:

$$SS\colon\ u = f(M, \sigma^2_{jht} = \Pr(z_{jht} = 1)).$$

A second approach for introducing soft constraints on the consideration set is accomplished by relating variables to the scale of the random utility error in a more general fashion. The idea behind variance reduction models is that smaller errors in a random utility model correspond to more deterministic consumer behavior, whereas larger errors correspond to behavior that is harder to explain or predict. Thus, offerings with a smaller error scale can be thought to comprise a set of brands that are considered more carefully, for which changes and prices and merchandising displays are more predictive of choice. Brands with a larger error scale are not as predictable, and as the scale increases, these brands can be viewed as not seriously considered (Allenby and Ginter 1995).

Consideration Sets via Variance Reduction

$$\text{V1:}\quad u = f(M, \sigma^2_{jht} = \sigma^2_j AS^{-1}_{jht});$$

$$u_{jht} = M'_{jht}\beta_h + \varepsilon_{jht};\quad \varepsilon_{jht} \sim \mathrm{N}(0, \sigma^2_{jht} = \sigma^2_j AS^{-1}_{jht}).$$

$$\text{V2:}\quad u = f(M, \sigma^2_{jht} = \sigma^2_j AS^{-2}_{jht});$$

$$u_{jht} = M'_{jht}\beta_h + \varepsilon_{jht};\quad \varepsilon_{jht} \sim \mathrm{N}(0, \sigma^2_{jht} = \sigma^2_j AS^{-2}_{jht}).$$

$$\text{V3:}\quad u = f(M, \sigma^2_{jht} = \sigma^2_j(1 + DS_{jht}) - 1);$$

$$u_{jht} = M'_{jht}\beta_h + \varepsilon_{jht};\quad \varepsilon_{jht} \sim \mathrm{N}(0, \sigma^2_{jht} = \sigma^2_{jh}(1 + DS_{jht})^{-1}).$$

$$\text{V4:}\quad u = f(M, \sigma^2_{jht} = \sigma^2_j(1 + DS^2_{jht})^{-1});$$

$$u_{jht} = M'_{jht}\beta_h + \varepsilon_{jht};\quad \varepsilon_{jht} \sim \mathrm{N}(0, \sigma^2_{jht} = \sigma^2_{jh}(1 + DS^2_{jht})^{-1}).$$

Finally, we make comparison to the model proposed by Erdem and Keane (1996, hereafter denoted as EK), where consumers learn about brand quality through advertising. We employ Equation (19) of EK and denote it as the consumer learning model CL:

$$CL\colon\ u = f\big(M, A_j, r_h, w_{\cdot h}, \sigma^2_{\cdot h}, \{\nu_{jh}(t)\}\big),$$

where the parameters r_h, $w_{\cdot h}$, $\sigma^2_{\cdot h} = \{\sigma^2_{\nu 0h}, \sigma^2_{bh}, \sigma^2_{\xi h}\}$, and $\{\nu_{jh}(t)\}$ are described by EK in their Equations (8), (11), (12), (14), and (15). Their model describes a process where advertising is incorporated into consumer perceptions of utility, with utility varying over time as new advertising is observed. We note that our model differs from EK by not incorporating the no-buy choice option, thus avoiding the need to assume the outside good is consistently defined over purchase occasions. This assumption is not plausible when consumer purchases reflect large expenditure trips and small expenditure trips used to purchase specific, out-of-stock goods. For the EK model, we use actual calendar time of advertising exposure and update consumer evaluations of utility at the time of purchase.

4. Results

Tables 2 and 3 report in-sample and predictive fits of the various models; Table 2 shows results for the detergent data and Table 3 shows results for the coffee data. In-sample fit is measured using the log-marginal density (LMD) of the data, a Bayesian statistic used to create Bayes factors for hypothesis testing. Larger (less negative) values of LMD indicate evidence for the model. Predictive fit is measured in two ways. The first is in terms of the hit rate, i.e., the proportion of time the chosen alternative has the largest expected utility, where the expectation is with respect to the posterior distribution of model parameters. The second predictive measure is the mean absolute deviation between the observed choice and the choice probability. Results for models S0–S9 and SS are broken down for whether the marketing variables, M, contain the brand loyalty covariate, and whether the threshold for consideration is brand specific.

The general ordering of the various models in terms of in-sample and predictive fit reveals that the proposed screening rule models fit data best (S0–S9), followed by the SS, variance reduction models (V1–V4), and direct effect models (N1–N6), including the consumer learning model (CL) of EK. The actual rank ordering is more mixed than this, but this general ordering follows from the best-fitting model within each category. We also find that allowing for brand-specific thresholds and including the brand loyalty covariates in M generally improve the model fit.

The fit statistics indicate support for the proposed model and indicate that the primary influence of media advertising is consideration set formation, not utility modification. This implication follows immediately from the fit statistics that indicate $\text{S0} \doteq \text{S1} \gg \text{N1} > \text{N0}$. Relating advertising stock to consideration set formation results in a large improvement in fit relative to models that allow advertising to directly modify the utility function. Allowing media advertising to also influence utility results in marginal improvements in fit beyond its effect on consideration set formation. This result is not sensitive to the inclusion of the brand loyalty covariate nor the brand-specific threshold.

We find that advertising stock results in more-accurate model fit statistics than past purchases (Fader and McAlister 1990) and that variance reduction models with display stock (V3 and V4) fit the data better than any of the direct effect models. We comment further on the similarities and differences of these models below. Finally, the consumer learning model of EK fits about the same as other models that assume advertising directly affects consumer utility of the brands.

Table 4 displays coefficient estimates for model S1 with common brand thresholds and no brand loyalty variable. The reason for selecting this model is

Table 2 **Model Fit for Laundry Detergent Data**

	Not containing *BL*						Containing *BL*					
	Category threshold			Brand threshold			Category threshold			Brand threshold		
Models	LMD	HR (holdout)	MAD (holdout)	LMD	HR (holdout)	MAD (holdout)	LMD	HR (holdout)	MAD (holdout)	LMD	HR (holdout)	MAD (holdout)
Direct effect models												
Linear models												
N0 $u = f(M)$	−4,636.87	41.136	0.2480				−3,161.38	48.217	0.2230			
N1 $u = f(M, AS)$	−4,161.34	48.145	0.2109				−4,161.34	48.145	0.2109			
N2 $u = f(M, AS, BL)$	−2,458.54	47.533	0.2159				−2,458.54	47.533	0.2159			
N3 $u = f(M, AS, BL, BL^2)$	−1,984.83	43.460	0.2245				−1,984.83	43.460	0.2245			
N4 $u = f(M, AS, BL, BL^2, BL^3)$	−2,182.69	40.905	0.2288				−2,182.69	40.905	0.2288			
Nonlinear models												
N5 $u = f(M, AS, AS \times \text{Price})$	−4,570.33	40.506	0.2185				−1,964.68	47.960	0.2435			
N6 $u = f(M, AS, AS^2)$	−3,554.15	43.847	0.2410				−2,255.45	45.686	0.2099			
N7 $u = f(M, AS, AS^2, AS^3)$	−3,996.99	45.002	0.1940				−2,926.63	41.581	0.2481			
Indirect effect models												
Screening models												
S0 $u = f(M, AS) \mid scr(AS)$	−1,212.02	59.060	0.1600	−1,112.06	58.741	0.1538	−1,190.40	59.119	0.1567	−1,006.09	60.249	0.1519
S1 $u = f(M) \mid scr(AS)$	−1,222.23	58.349	0.1603	−1,128.56	59.155	0.1551	−1,209.17	58.811	0.1555	−1,042.76	60.054	0.1480
S2 $u = f(M) \mid scr(BL)$	−1,277.33	57.326	0.1669	−1,136.01	56.473	0.1714	−1,286.87	58.660	0.1609	−1,045.08	56.243	0.1606
S3 $u = f(M) \mid scr(DS)$	−1,300.57	58.117	0.1614	−1,240.22	58.902	0.1565	−1,266.10	55.249	0.1648	−1,148.49	59.204	0.1516
S4 $u = f(M) \mid (scr(AS)$ or $scr(BL))$	−1,385.45	56.938	0.1616	−1,340.12	57.177	0.1620	−1,310.66	57.347	0.1713	−1,220.52	57.433	0.1693
S5 $u = f(M) \mid (scr(AS)$ or $scr(DS))$	−1,392.64	58.010	0.1612	−1,305.69	58.634	0.1568	−1,289.88	57.057	0.1623	−1,192.85	58.591	0.1548
S6 $u = f(M) \mid (scr(AS)$ or $scr(BL)$ or $scr(DS))$	−1,395.96	58.285	0.1607	−1,316.27	57.907	0.1590	−1,294.97	57.136	0.1615	−1,179.88	58.520	0.1538
S7 $u = f(M) \mid (scr(AS)$ and $scr(BL))$	−1,348.89	56.955	0.1641	−1,421.89	53.660	0.1636	−1,425.92	57.683	0.1694	−1,321.32	55.795	0.1742
S8 $u = f(M) \mid (scr(AS)$ and $scr(DS))$	−1,327.16	57.772	0.1677	−1,413.97	58.160	0.1625	−1,352.73	56.303	0.1708	−1,306.99	56.981	0.1541
S9 $u = f(M) \mid (scr(AS)$ and $scr(BL)$ and $scr(DS))$	−1,332.74	57.241	0.1674	−1,426.71	55.915	0.1661	−1,429.26	54.080	0.1693	−1,326.25	54.616	0.1567
Stochastic screening model												
SS $u = f(M, \sigma_{jht}^2 = \Pr\{AS_{jht}\}^{-1})$	−1,969.52	49.801	0.1734	−1,808.64	51.295	0.1635	−1,965.53	52.618	0.1747	−1,829.20	52.950	0.1620
Variance reduction model												
V1 $u = f(M, \sigma_{jht}^2 = \sigma_j^2 AS_{jht}^{-1})$	−2,545.93	33.873	0.1825				−3,080.88	38.465	0.1803			
V2 $u = f(M, \sigma_{jht}^2 = \sigma_j^2 AS_{jht}^{-2})$	−3,455.21	28.860	0.1900				−3,913.65	35.504	0.2044			
V3 $u = f(M, \sigma_{jht}^2 = \sigma_j^2 (1 + DS_{jht})^{-1})$	−1,970.68	52.814	0.1765				−2,129.28	48.765	0.1767			
V4 $u = f(M, \sigma_{jht}^2 = \sigma_j^2 (1 + DS_{jht}^2)^{-1})$	−1,939.01	53.411	0.1744				−2,124.22	46.624	0.1776			
EK model	−2,636.13	43.670	0.1712				−2,208.21	46.803	0.1765			

Note. HR and MAD stand for hit rate and mean absolute deviation, respectively, for the holdout sample.

Table 3 Model Fit for Instant Coffee Data

	Not containing *BL*						Containing *BL*					
	Category threshold			Brand threshold			Category threshold			Brand threshold		
Models	LMD	HR (holdout)	MAD (holdout)	LMD	HR (holdout)	MAD (holdout)	LMD	HR (holdout)	MAD (holdout)	LMD	HR (holdout)	MAD (holdout)
Direct effect models												
Linear models												
N0 $u=f(M)$	−3,194.79	46.232	0.2648				−2,586.59	49.611	0.2448			
N1 $u=f(M, AS)$	−2,992.78	48.588	0.2329				−2,992.78	48.588	0.2329			
N2 $u=f(M, AS, BL)$	−2,653.01	62.183	0.2242				−2,653.01	62.183	0.2242			
N3 $u=f(M, AS, BL, BL^2)$	−2,633.33	62.667	0.2278				−2,633.33	62.667	0.2278			
N4 $u=f(M, AS, BL, BL^2, BL^3)$	−2,817.88	59.530	0.2397				−2,817.88	59.530	0.2397			
Nonlinear models												
N5 $u=f(M, AS, AS\times\text{Price})$	−3,770.37	44.505	0.3291				−2,631.58	45.574	0.2839			
N6 $u=f(M, AS, AS^2)$	−2,821.68	55.602	0.2087				−2,498.85	44.650	0.2294			
N7 $u=f(M, AS, AS^2, AS^3)$	−2,893.47	45.519	0.3270				−2,624.43	47.155	0.2582			
Indirect effect models												
Screening models												
S0 $u=f(M, AS) \mid scr(AS)$	−476.05	80.042	0.0503	−462.10	80.913	0.0501	−457.60	78.477	0.0559	−444.42	81.588	0.0511
S1 $u=f(M) \mid scr(AS)$	−470.99	77.529	0.0502	−464.64	78.075	0.0492	−452.45	73.698	0.0507	−439.98	76.143	0.0447
S2 $u=f(M) \mid scr(BL)$	−925.71	64.032	0.1730	−640.24	61.270	0.1284	−674.47	67.152	0.1751	−442.35	69.905	0.1832
S3 $u=f(M) \mid scr(DS)$	−486.89	75.613	0.0496	−469.30	76.256	0.0481	−534.71	69.713	0.0459	−518.87	74.080	0.0483
S4 $u=f(M) \mid (scr(AS)$ or $scr(BL))$	−521.40	72.050	0.1593	−479.32	73.323	0.1136	−506.70	72.627	0.1353	−447.87	70.468	0.1576
S5 $u=f(M) \mid (scr(AS)$ or $scr(DS))$	−492.14	74.808	0.0511	−484.74	75.845	0.0515	−481.17	72.028	0.0635	−465.52	70.226	0.0492
S6 $u=f(M) \mid (scr(AS)$ or $scr(BL)$ or $scr(DS))$	−492.90	74.310	0.0705	−482.67	72.754	0.0802	−474.14	72.558	0.1093	−447.27	73.310	0.0976
S7 $u=f(M) \mid (scr(AS)$ and $scr(BL))$	−1,027.75	68.087	0.1002	−895.21	68.785	0.1079	−953.57	67.849	0.1076	−599.24	67.548	0.1050
S8 $u=f(M) \mid (scr(AS)$ and $scr(DS))$	−475.64	79.025	0.0548	−586.62	76.230	0.0603	−467.69	61.878	0.0776	−691.70	73.549	0.0651
S9 $u=f(M) \mid (scr(AS)$ and $scr(BL)$ and $scr(DS))$	−915.57	69.893	0.1425	−596.79	72.653	0.1306	−945.37	70.304	0.1720	−1,057.35	73.324	0.1291
Stochastic screening model												
SS $u=f(M, \sigma_{jht}^2=\Pr\{AS_{jht}\}^{-1})$	−2,512.17	71.562	0.1337	−1,561.48	72.631	0.1087	−1,889.53	72.418	0.1360	−1,550.36	72.842	0.1182
Variance reduction model												
V1 $u=f(M, \sigma_{jht}^2=\sigma_j^2 AS_{jht}^{-1})$	−5,751.68	51.036	0.4255				−3,297.54	51.610	0.3944			
V2 $u=f(M, \sigma_{jht}^2=\sigma_j^2 AS_{jht}^{-2})$	−5,681.09	47.217	0.4270				−3,583.92	54.151	0.3524			
V3 $u=f(M, \sigma_{jht}^2=\sigma_j^2(1+DS_{jht})^{-1})$	−1,933.96	60.177	0.1816				−1,912.27	54.426	0.2415			
V4 $u=f(M, \sigma_{jht}^2=\sigma_j^2(1+DS_{jht}^2)^{-1})$	−1,941.49	60.003	0.1807				−2,020.39	53.160	0.2116			
EK model	−2,403.25	45.754	0.2609				−1,789.81	45.756	0.2172			

Note. HR and MAD stand for hit rate and mean absolute deviation, respectively, for the holdout sample.

Table 4 Posterior Means of Heterogeneity Distribution

Parameter	S1 ($u = f(M) \mid scr(AS)$)	S1 ($u = f(M) \mid scr(AS)$)
	Laundry detergent	Instant coffee
Threshold level		
AS	1.056	1.456
	(1.421)	(1.004)
Market response parameter		
Display	0.062	20.010
	(0.265)	(0.054)
Price	−4.787	−1.955
	(0.239)	(0.108)
Constant—Brand 1	−0.367	−0.179
	(0.705)	(0.742)
Constant—Brand 2	−0.853	−2.554
	(0.689)	(1.353)
Constant—Brand 3	−0.486	−2.191
	(0.737)	(1.224)
Constant—Brand 4	−0.286	N/A
	(0.750)	
Constant—Brand 5	−1.033	N/A
	(0.788)	
Carryover parameter		
Advertising carryover	0.905	0.821
	(0.105)	(0.145)

Note. Standard deviations of heterogeneity are in parentheses.

that whereas the inclusion of brand-specific thresholds improves the in-sample fit of the data, the predictive fit often favors the simpler model with a common threshold. Similarly, the models with brand loyalty included in the set of covariates, *M*, do not lead to uniform improvement in predictive results. The parameter estimates, however, are generally similar across the model specifications and are available from the authors. We comment on the differences of estimated coefficients across models below.

Reported are posterior means of the random effects distribution and the standard deviation of heterogeneity. The parameters are estimated with reasonable algebraic signs (e.g., price coefficients are negative), and the brand intercepts agree with the ordering implied by the choice shares reported in Table 1. For the coffee data, we note that the display coefficients are very large. The reason for this is due to the merchandising behavior in this category—i.e., displays rarely occur (see Table 1), and when they do, they are never present for multiple brands. Moreover, we find that a brand was selected whenever it was displayed, accounting for approximately 10% of observed purchases. Although we believe this leads to abnormally large display coefficients for coffee, it does not necessarily affect the other coefficients in the model that are also influenced by the other 90% of the data. This result is found across all model specifications.

The carryover coefficients for *AS* are reported at the bottom of Table 3 and are large. For detergent, the average estimate is 0.905, whereas for coffee, the average estimate is 0.821. These estimates indicate that, on average, advertising stock is diminished from about 10%–20% each observational interval. In models with brand-specific threshold, these estimates reduce to about 0.70. It is important to remember that these estimates pertain to the average of population because the carryover coefficient is modeled with heterogeneity. A posterior 95% credible interval for the implied monthly decay rates can be obtained by first identifying the 95% credible interval for the estimated distribution of heterogeneity and then transforming the endpoints by raising them to the appropriate power to reflect the difference in timescale—i.e., 4 for the weekly detergent data and 30 for the coffee data. The implied 95% heterogeneity interval of carryover effects is (0.15, 0.98) for detergent and (0.00, 0.71) for coffee. Thus, monthly carryover is nonnegligent in our analysis (Tellis 2004).

Finally, the threshold coefficient is reported at the top of Table 4. The estimated mean value of the distribution of heterogeneity for this parameter in model S1 is 1.056 for the detergent data, with the standard deviation of heterogeneity equal to 1.421. As discussed in the appendix, this parameter is constrained to be positive, implying that the distribution is skewed with a right tail. As the threshold parameters approach 0, or the carryover coefficient approaches 1, the consideration set converges to the entire choice set. Thus, the proposed models nest a standard choice model with no consideration component and that has the ability to reflect no consideration choices if needed.

Table 5 reports the consideration set inclusion rates for brands in both categories for each of the screening models S0–S9. We find general agreement among the models for the high-share brands and differences for the low-share brands—brands 4 and 5 in detergent and brands 2 and 3 in coffee (see Table 1). The rates of consideration set inclusion are different for models S1 and S2, where the consideration set for S2 is based on past purchases (*BL*) instead of advertising stock (*AS*). When advertising stock is used to form consideration sets, brands with greater advertising activity are present in the consideration set more often, differentially affecting the low-share brands. Whereas the aggregate fit statistics of models S1 and S2 for detergent are similar, their implications for choice set consideration are different for brands that are infrequently purchased. In contrast, advertising activity is vastly different for low-share brands in coffee, and the fit statistics for models S1 and S2 reflect this difference. It appears that the use of brand loyalty, or state dependence, works as a screening variable when past purchases are reasonable surrogate measures of advertising exposure. This is not the case for the coffee data, where the fit statistics favor consideration

Table 5 Choice Set Inclusion

Laundry detergent data	Brand 1 (%)	Brand 2 (%)	Brand 3 (%)	Brand 4 (%)	Brand 5 (%)	Brand 6 (%)
S0 $u = f(M, AS) \mid scr(AS)$	92.7	94.2	79.5	58.1	87.2	98.8
S1 $u = f(M) \mid scr(AS)$	93.8	95.2	77.6	55.6	87.2	99.1
S2 $u = f(M) \mid scr(BL)$	88.9	80.1	77.6	72.7	72.7	97.4
S3 $u = f(M) \mid scr(DS)$	97.8	85.5	83.3	96.4	89.3	99.7
S4 $u = f(M) \mid (scr(AS)$ or $scr(BL))$	98.6	98.9	94.6	89.7	97.0	99.7
S5 $u = f(M) \mid (scr(AS)$ or $scr(DS))$	99.8	99.3	94.6	98.5	99.3	100.0
S6 $u = f(M) \mid (scr(AS)$ or $scr(BL)$ or $scr(DS))$	100.0	99.6	98.3	99.9	100.0	100.0
S7 $u = f(M) \mid (scr(AS)$ and $scr(BL))$	85.6	83.7	72.8	53.6	73.2	97.0
S8 $u = f(M) \mid (scr(AS)$ and $scr(DS))$	90.3	83.6	74.0	60.5	79.2	98.8
S9 $u = f(M) \mid (scr(AS)$ and $scr(BL)$ and $scr(DS))$	84.7	76.3	67.4	53.6	69.5	96.9

Instant coffee data	Brand 1 (%)	Brand 2 (%)	Brand 3 (%)	Brand 4 (%)
S0 $u = f(M, AS) \mid scr(AS)$	96.9	94.1	62.6	51.9
S1 $u = f(M) \mid scr(AS)$	97.6	94.6	64.3	53.1
S2 $u = f(M) \mid scr(BL)$	80.8	50.3	47.9	54.9
S3 $u = f(M) \mid scr(DS)$	97.9	90.4	90.8	92.0
S4 $u = f(M) \mid (scr(AS)$ or $scr(BL))$	98.2	95.0	75.1	68.2
S5 $u = f(M) \mid (scr(AS)$ or $scr(DS))$	99.9	98.6	95.1	94.1
S6 $u = f(M) \mid (scr(AS)$ or $scr(BL)$ or $scr(DS))$	99.9	98.8	95.8	94.3
S7 $u = f(M) \mid (scr(AS)$ and $scr(BL))$	80.8	51.6	38.5	42.7
S8 $u = f(M) \mid (scr(AS)$ and $scr(DS))$	96.4	86.4	58.5	51.2
S9 $u = f(M) \mid (scr(AS)$ and $scr(BL)$ and $scr(DS))$	80.9	50.9	38.5	43.3

Note. For laundry detergent data, rates based on 730 purchase occasions; for instant coffee data, rates based on 551 purchase occasions.

set formation based on the *AS* variable rather than *BL* variable.

5. Discussion

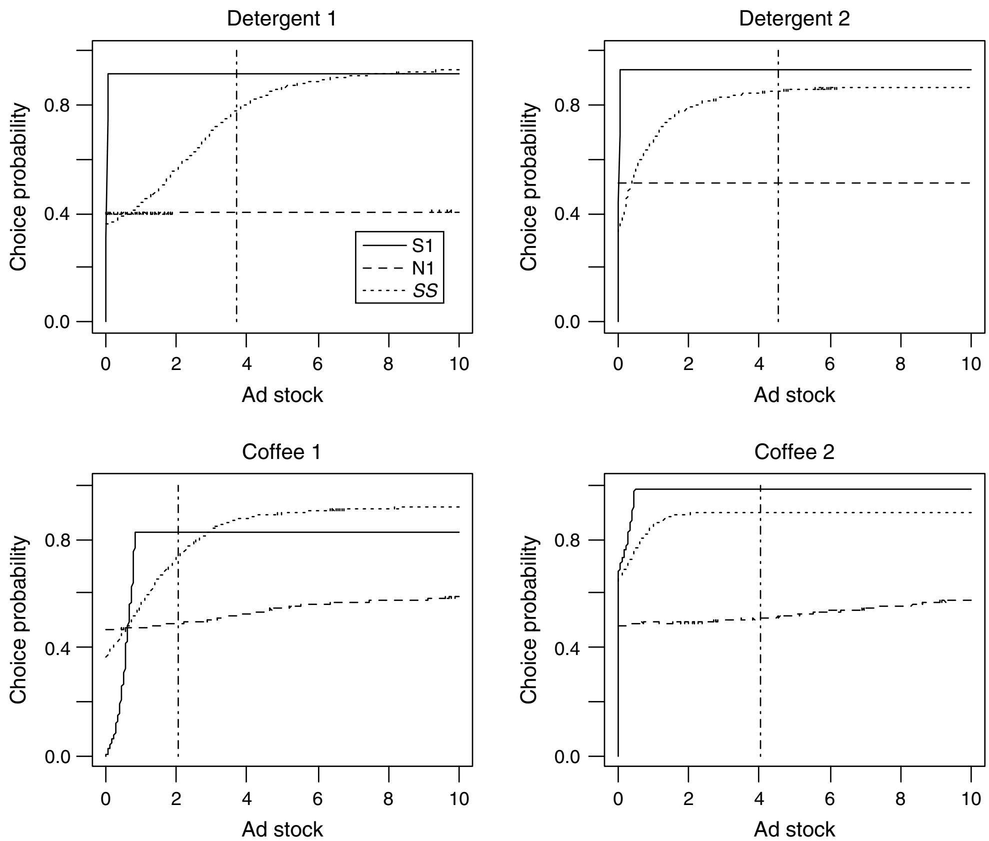
The results indicate that the effect of media advertising on brand choice is primarily mediated through the construction of respondent consideration sets. This finding is at odds with much of the existing empirical literature on media advertising, where effects are thought to interact with other variables such as prices or exhibit nonlinear effects through the presence of thresholds or consumer learning about product attributes (Erdem et al. 2008). The fit statistics reported in Tables 2 and 3 provide strong evidence that consideration sets, and media advertising's role in their formation, provide a better explanation of purchase data in the laundry detergent and instant coffee categories.

Two explanations have been offered for why this effect is present (Chakravarti et al. 2006). The first explanation is that the mere act of screening encourages respondents to deemphasize information used to screen in subsequent choice and to emphasize information not used. That is, screening information has already been used and is not used again. The second explanation is that screening serves to produce a considered set of alternatives that are relatively homogeneous with regard to the screening variable. In our model of consideration, all that matters is that the advertising stock is above a threshold value, rendering the screened choices identical with regard to being above the threshold. If the effect of advertising is, in fact, dichotomous, then it will not play a role in choices among the considered brands. Our model fit statistics indicate strong support for advertising having a dichotomous effect.

The large improvement in model fit for the proposed model is due to the hard constraint on brand inclusion in the consideration set. Figure 2 provides a comparison of the detergent and coffee brand choice probability as a function of the advertising stock variable for the best-fitting model S1, the stochastic screening rule model (*SS*), and a standard choice model (N1). Plotted are the estimated choice probabilities for four randomly selected households, each for a randomly selected choice occasion. The choice probabilities are evaluated by integrating the likelihood of choice with respect to the posterior distribution of model parameters. In addition to the probability profiles for the three models, Figure 2 includes a vertical line indicating the current estimated advertising stock for that choice occasion.

The choice probability profiles indicate that the proposed model (S1) has a profile resembling a step function, with low levels of advertising stock leading to an abrupt decrease in the choice probability of low values of advertising stock. At higher levels of magnification, and more generally, for the instant coffee data, the profile for the S1 model is less abrupt. A brand's absence from the consideration set is associated with a zero choice probability in the S1 model, and this

Figure 2 Choice Probabilities and Advertising Stock



probability mass is redistributed among the brands in the consideration set. The result is a higher purchase probability for brands in the consideration set, with the solid line for model S1 taking on higher values than the dashed (N1) or dotted (*SS*) profiles. Across brands and observations, this leads to the higher in-sample and predictive fit reported in Tables 2 and 3.

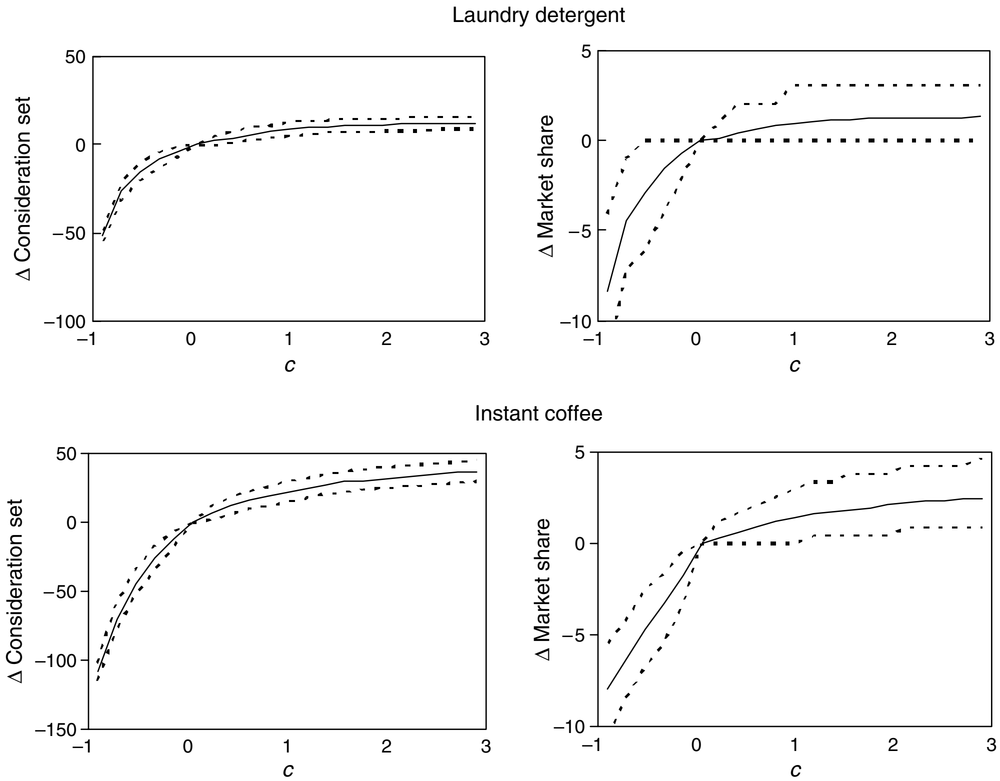
Many stochastic screening rule models associate separate error terms for consideration and choice (see van Nierop et al. 2010). These models have two errors associated with each observation, which are not uniquely identified without additional constraints or the use of temporal information. Theory offers little guidance for allocating the error variances. In addition, we show in the appendix that these models can inflate the overall model error when unit scale values are assumed as in standard logit and probit models. Such inflation and arbitrary allocation leads to a decreased fit of stochastic screening rule models to the data. We believe our model with hard constraints on the consideration set more truly represents the concept of consideration—brands are either considered or they are not, and the mechanism for consideration is not clouded by the presence of additional error terms.

Implications

We highlight two implications of the presence of media-driven consideration sets for marketing. The first is that, for mature packaged products like those studied in our analysis, our model structure implies that one advertising exposure is sufficient to move a brand into a consumer's consideration set. This is because our model has sharp boundaries on brand inclusion, as opposed to soft boundaries as in the stochastic screening rule model of Bronnenberg and Vanhonaker (1996) and other variance reduction models (V1–V4). Sharp boundaries imply that the effect of advertising occurs at a much quicker rate than generally expected, where three "hits" are often used as a norm for achieving an advertising effect (Krugman 1972). The implication is that advertising reach, or the percentage of a target audience that is exposed at least once to a message, is a more appropriate measure of advertising exposure than gross rating points (GRPs). GRPs are measured as the product of reach and the frequency of exposure, and they are given too much weight to multiple exposures when just one advertising impression is sufficient for a brand being included in a household's consideration set.

A second implication of our model is investigated by considering the effect of proportional changes to

Figure 3 **Asymmetric Advertising Effects for Brand 3**



current advertising rates. This is accomplished by modifying the observed advertising data as

$$a^*_{jht} = (1+c)a_{jht}, \tag{8}$$

where a_{jht} is a variable indicating that household h has viewed an advertisement for brand j at time t, and a value of $c = 0.2$, for example, can be interpreted as a 20% increase in advertising. The modified viewing data take on values other than 0 and 1, and they can be interpreted as viewing multiple (or fractions) of advertisement during the weeks that an advertisements was originally viewed. Because values of a_{jht} equal to 0 are not affected by the value of the constant c, the timing of advertising insertions is not affected.

The effect of changing the advertising viewing data for different values of c is displayed in Figure 3 for brand 3 in the detergent and coffee data. Similar results are found for all brands in our analysis. Plotted is the expected change in the number of times brand 3 is included in the consideration set and the change in the expected choice share of the brand. Error bounds corresponding to a 95% credible interval are also included. The predictions are obtained by integrating over the posterior distribution of household parameters and the observed data, modified by different values of the constant c that are plotted along the horizontal axis.

We find that the effects of advertising are asymmetric, with declines in advertising ($c < 0$) having a greater effect on consideration set inclusion and market share than advertising gains ($c > 0$). Brands have to be in the consideration set for a purchase to occur, but once this occurs a purchase only results if the brand affords the consumer the greatest utility among the alternatives. The presence of a consideration set provides an alternative explanation for asymmetry than, for example, prospect theory (Khaneman and Tversky 1979) or nonhomothetic utility (Allenby et al. 2010).

6. Concluding Remarks

This paper presents a controversial finding—advertising primarily affects brand consideration, not brand utility, in two frequently purchased product categories. Although it is premature to generalize our results beyond the types of products studied in our analysis—i.e., mature packaged goods—and we acknowledge the importance of consumer learning for new products and products with nontangible attributes, our results indicate support for the idea that media advertising information is used once, not twice, in determining brand choice. That is, in the context of mature packaged goods like instant coffee and laundry detergent, it affects brand consideration but not brand utility. Moreover, we find support for the notion that advertising has a dichotomous effect—i.e., brands either come to mind or they do not, and after a certain level of advertising exposure, additional advertising does not help. Our results indicate that satiation effects for advertising occur abruptly.

There are qualifications and potential objections that need to be addressed to limit the scope of our findings. There are many situations where we expect media advertising to have a direct effect on consumer learning and consumer sensitivity to the presence and enhancement of product attributes. One obvious area is that of new product introductions and studies associated with brand re-formation. Here, advertising is expected to play an important role in informing and persuading consumers. Another is the announcement of short-term price promotions, but for mature packaged products, such as those examined in our analysis (coffee and detergent), where consumers are already aware of the brands whose formulation is fairly consistent over time, we believe our results will hold up to replication in other product categories—i.e., general brand advertising on the television does not affect the utility of frequently purchased mature packaged products.

We have discussed previously the issue of endogeneity and reasons why we do not believe this is an issue in our data because of the large lead time in purchasing advertising. However, we acknowledge that we have not provided direct evidence that successfully modeling endogeneity does not change our empirical result. Additional work is needed to more explicitly investigate this potential effect and others, such as in exploring the role of merchandising variables in consideration set formation and modeling the amount of television viewing by respondents. We find that display variables act to focus consumer attention on specific brands (models V3 and V4) but do not act in a manner consistent with a strict consideration set where brands not included in the set have a zero likelihood of purchase. Thus, brands not displayed at all are still purchased. The evidence presented in our analysis, which agrees with previously reported studies, is that brand display variables increase the likelihood of purchase while reducing the presence of competitive effects. In addition, our model assumes an identity covariance matrix of choice errors, which could be improved upon by assuming a general covariance matrix. Finally, although we present analyses and comparisons across a number of different model specifications, we have certainly not exhausted the set of possible alternative model structures. Examples include a more general exponential smoothing process and possibly a more general "grand" model that allows for advertising to exhibit itself through alternative algebraic expressions that affect utility, carryover, and consideration set formation. Further work is needed to explicitly model these combined effects.

Appendix

Model Formulation

Given the model parameters ρ_h, advertising stock variables are constructed for each brand (j) and time period using the recursive relationship:

$$AS_{jhw} = a_{jhw} + \rho_h AS_{jh,\,w-1} \quad (0 \le \rho_h < 1), \tag{9}$$

where w denotes calendar time, a_{jhw} is the number of advertising exposures of brand j to household h at time w, and ρ_h is a heterogeneous carryover parameter. We note that the calendar time w is not always related to purchase time, denoted as t. Other stock variables examined in our analysis are constructed similarly.

Following Erdem et al. (2008) and Keane (1997), we set the initial value of advertising stock, and other stock variables, as $AS_{jh_0} = (1/(1-\rho_h))(\sum_{w=1}^{W} a_{jhw})/W$, where W is the total number of periods (in calendar time) used in estimating model parameters. Model results are not expected to be sensitive to the value of initial stock for brands in mature packaged product categories, such as those we study in the following.

Purchase history data comprising observed choices $\{y_{jht}\}$ and corresponding marketing variables $\{x_{jht}\}$ are denoted $\text{PData}_{ht} = (y_{jht}, x_{jht}, j = 1, \ldots, N)$ for $t = 1, \ldots, T_h$. In addition to household purchases collected through conventional loyalty cards, we also employ advertising exposure data collected from an "in-home" viewing device connected to a respondent's television set. We denote advertising data for the observation period $t = 1, \ldots, W$ as $\{\text{AData}_{ht}\}$. To express the dependence of advertising stock on carryover parameter expressed in (4), we denote the advertising stock as $AS_{jhw}(\rho_h)$. Then, using an advertising screening rule to PData_{ht} at time t, we obtain the relative subset of purchase data at time t corresponding to the brands included in the consideration set $\text{PData}_{ht}^{C_{ht}^{[AS_{jht} \ge r_h]}}$ $(\subseteq \text{PData}_{ht})$. The conditional likelihood for β_h, given $\phi_h = (r_h, \rho_h)$, is then

$$\text{L}(\{\beta_h\} \mid \{\text{PData}_{ht}\}, \{\text{AData}_{ht}\}, \{\phi_h\}) \tag{10}$$

$$= \prod_{h=1}^{H} \prod_{t=1}^{T_h} \text{L}(\beta_h \mid \text{PData}_{ht}, \text{AData}_{ht}, \phi_h) \tag{11}$$

$$= \prod_{h=1}^{H} \prod_{t=1}^{T_h} \left[\prod (p_{jht})^{y_{jht}} \colon\; j \in C_{ht}^{[AS_{jht}(\rho_h) \ge r_h]} \right], \tag{12}$$

where $p_{jht} = \Pr\{u_{jt} = \max\{u_{kht} \colon k \in C_{ht}^{[AS_{jht}(\rho_h) \ge r_h]}\}\}$.

Bayesian data augmentation is used to avoid direct evaluation of the choice probabilities in Equation (10). For brands passing the screen, latent utilities are generated from a truncated normal distribution (see Rossi et al. 2005). For brands not passing the screen, latent utilities are not drawn and are not used in to make inferences about β_h. This is reasonable in that the (indirect) utility function in (2) includes brand price, which must be combined with model intercepts to arrive at the latent value used in estimation. In our model, we assume that this calculation is not performed for brands not considered.

Our estimation approach is similar to Gilbride and Allenby (2004) but differs from their algorithm because of the presence of two sets of data (advertising exposure and brand purchase) where consideration set formation is determined by data and parameters that are exogenous to the data and parameters of conditional choice. In addition, our screening rules are dynamic over time, depending on the level of the stock variable.

The conditional likelihoods for the threshold and carryover parameters, e.g., $\phi_h = \{\rho_h, r_h\}$, given the latent utilities u_{jht} and coefficients β_h take the form of truncated distributions because of restrictions needed for logical consistency. A chosen alternative needs to be present in the consideration set for it to be chosen, implying that its advertising stock AS_{kht} is greater than the threshold r_h. This imposes a restriction on the augmented parameter space as

$$R_{\phi_h} = \{\phi_h = (r_h, \rho_h)\colon [AS_{kht}(\rho_h) \geq r_h \geq 0] \tag{13}$$

$$\text{for all } t = 1, \ldots, T_h, \text{ where } y_{kht} = 1\}. \tag{14}$$

Finally, we specify parameters heterogeneously using standard hierarchical random-effect models:

$$r_h \sim \mathrm{N}(\bar{r}, \nu_r), \tag{15}$$

$$\rho_h^* = \ln \frac{\rho_h}{1-\rho_h} \sim \mathrm{N}(\bar{\rho}^*, \xi_\rho) \quad (0 \leq \rho_h < 1), \tag{16}$$

$$\beta_h \sim \mathrm{N}(\bar{\beta}, \Lambda). \tag{17}$$

Estimation proceeds by recursively generating draws from the full conditional distributions of model parameters using the Metropolis-Hastings algorithm when conjugate conditional distributions are not available. A simulation study, available from the authors, confirms the convergence of the algorithm to the posterior distribution of model parameters from multiple starting conditions.

Our data set does not contain household demographic information that could be included in the distribution of heterogeneity. We therefore employ random effects models without covariates in our analysis, treating all heterogeneity as unobservable. It is also important to note that our model is a conditional demand model and does not attempt to model the no-purchase decision. Thus, we do not consider effects of physical inventory on purchase timing. The priors on hyperparameters and their induced conditional posterior distributions are described below.

Stochastic Screening Rule Model. The conditional likelihood, when the state z_{jht} of brand j is given, is expressed as

$$\mathrm{L}(\{\beta_h\}, \{\phi_h\}; \{\mathrm{PData}_{ht}\} \mid \{\mathrm{AData}_{ht}\}, \{z_{jht}\}) \tag{18}$$

$$\propto \prod_{h=1}^{H} \prod_{t=1}^{T_h} \prod_{j=1}^{N} \exp\{-\tfrac{1}{2}(u_{jht} - X_{jht}\beta_h)^2 z_{jht}\}. \tag{19}$$

However, we do not know the true state $\{z_{jht}\}$, so we replace it by the inclusion probability $\{p_{1jht}\}$. Then we have the conditional likelihood on $\theta_h = (\phi_h, \delta_h)$ as

$$\mathrm{L}(\{\beta_h\}; \{\mathrm{PData}_{ht}\} \mid \{\mathrm{AData}_{ht}\}, \{p_{1jht}(\theta_h)\}) \tag{20}$$

$$\propto \prod_{h=1}^{H} \prod_{t=1}^{T_h} \prod_{j=1}^{N} \exp\{-\tfrac{1}{2}(u_{jht} - X_{jht}\beta_h)^2 p_{1jht}(\theta_h)\}, \tag{21}$$

where

$$u_{jht} \sim \mathrm{N}(X_{jht}\beta_h, p_{1jht}(\theta_h)^{-1}). \tag{22}$$

The conditional likelihood of θ_h is given by

$$\mathrm{L}(\{\theta_h\}; \{\mathrm{PData}_{ht}\} \mid \{\beta_h\}, \{\mathrm{AData}_{ht}\}) \tag{23}$$

$$\propto \prod_{h=1}^{H} \prod_{t=1}^{T_h} \prod_{j=1}^{N} \exp\{-c_{jht} p_{1jht}(\theta_h)\} \tag{24}$$

$$= \prod_{h=1}^{H} \prod_{t=1}^{T_h} \prod_{j=1}^{N} \exp\{-c_{jht}[1 + \exp\{-\delta_h(AS_{jht}(\rho_h) - r_h)\}]^{-1}\}, \tag{25}$$

where $c_{jht} = \frac{1}{2}(u_{jht} - X_{jht}\beta_h)^2$. We note that, in this model, there is no restriction over augmented parameter space on ϕ_h as in (13).

Markov Chain Monte Carlo Estimation

Model estimation proceeds by sequentially generating draws from the following distributions. Prior distributions for all model parameters are listed at the end of this section.

1. Latent Utility: $u_{jht} \mid \beta_h, \phi_h, \{\mathrm{Data}_{ht}\}$.

$$\text{If } S_{jht}(\rho_h) \geq r_h, \text{ e.g., } j \in C_{ht}(\phi_h),$$

(i) if $y_{jht} = 1$,

$$u_{jht} \sim TN(x'_{jht}\beta_h\colon U_{jht} > u_{kht} \text{ for all } k \in C_{ht}(\phi_h));$$

(ii) if $y_{jht} = 0$,

$$u_{jht} \sim TN(x'_{jht}\beta_h\colon u_{jht} \leq u_{kht} \text{ for all } k \in C_{ht}(\phi_h));$$

and we set $\mathrm{PData}_{ht}^{[C_{ht}(\phi_h)]} = \{(y_{jht}; u_{jht}, x_{jht}) \text{ for } j \in C_{ht}(\phi_h)\}$; otherwise, we do not generate any utilities.

2. Market Response and Its Covariance Parameters:
(2-1) $\beta_h \mid \phi_h, \Lambda, \bar{\beta}, \{\mathrm{Data}_{ht}\}$;

$$\beta_h \sim \mathrm{N}(\tilde{\beta}, (X_h(\phi_h)' X_h(\phi_h) + \Lambda^{-1})^{-1});$$

$$\tilde{\beta} = (X_h(\phi_h)' X_h(\phi_h) + \Lambda^{-1})^{-1} (X_h(\phi_h)' U_h + \Lambda^{-1}\bar{\beta}).$$

(2-2) $\bar{\beta} \mid \beta_h, \phi_h, \Lambda, \{\mathrm{Data}_{ht}\}$.
(2-3) $\Lambda \mid \beta_h, \phi_h, \bar{\beta}, \{\mathrm{Data}_{ht}\}$.

(2-2) and (2-3) are, respectively, a Normal distribution and an inverted Wishart (IW) distribution, which are available in Rossi et al. (1996, pp. 338–339).

3. Augmented Parameter: $\phi_h \mid \beta_h, (\bar{r}, \nu_r), (\bar{\rho}^*, \xi_\rho)\{\mathrm{Data}_{ht}\}$.

As for parameters $\phi_h = (r_h, \rho_h)$, we first obtain the conditional likelihood for each parameter by conditioning on ρ_h and r_h alternately, and it is combined with the prior to derive the conditional posterior distribution.

(3-1) Threshold Parameter: $r_h \mid \beta_h, \rho_h, (\bar{r}, \nu_r)$ $\{\mathrm{Data}_{ht}\}$.

Interchanging β_h and r_h in (10) leads to the conditional likelihood $\mathrm{L}(\{r_h\};\{\mathrm{PData}_{ht}\} \mid \{\rho_h\}, \{\beta_h\}, \{\mathrm{AData}_{ht}\})$, and it is combined with prior density (15) to get the posterior density of r as

$$\pi(r_h \mid \beta_h, \rho_h, \bar{r}, \nu_r, \{\mathrm{Data}_{ht}\})$$
$$\propto P(r_h \mid \bar{r}, \nu_r)\, \mathrm{L}(\{r_h\}; \{\mathrm{PData}_{ht}\} \mid \{\rho_h\}, \{\beta_h\}\{\mathrm{AData}_{ht}\}).$$

A Metropolis-Hastings algorithm with the restriction is applied to draw

$$r_h^{(d)} \sim TN\Big(r_h^{(d-1)}\colon \min_t S_{kht}(\rho_h) \geq r_h^{(d)} \geq 0, \text{ where } y_{kht} = 1\Big).$$

Then the acceptance probability α is defined as

$$\alpha\big(r_h^{(d)}, r_h^{(d-1)} \mid \beta_h, \rho_h, \bar{r}, \nu_r, -\big) = \min\left[\frac{\pi(r_h^{(d)} \mid \beta_h, \rho_h, \bar{r}, \nu_r, -)}{\pi(r_h^{(d-1)} \mid \beta_h, \rho_h, \bar{r}, \nu_r, -)}, 1\right].$$

(3-2) Carryover Parameter: $\rho_h \mid \beta_h, r_h, \bar{\rho}^*, \xi_\rho, \{\text{Data}_{ht}\}$.

First, the posterior density of reparameterized ρ_h^* in (13) can be formulated as

$$\begin{aligned}&\pi\big(\rho_h^* \mid \beta_h, r_h, \bar{\rho}^*, \xi_\rho, \{\text{Data}_{ht}\}\big)\\ &\quad\propto P\big(\rho_h^* \mid \bar{\rho}^*, \xi_\rho\big)\\ &\quad\text{L}\big(\{\rho_h\}; \{\text{PData}_{ht}\} \mid \{r_h\}, \{\beta_h\}\{\text{AData}_{ht}\}\big)\big|J_{\rho->\rho^*}\big|,\end{aligned}$$

where $J_{\rho\to\rho^*}$ is the Jacobian of transformation. A Metropolis-Hastings random walk algorithm for ρ_h^* is used for the sampling:

$$\rho_h^*(d) = \rho_h^{*(d-1)} + \omega_{\rho h}, \quad \omega_{\rho h} \sim \text{N}(0, 0.01),$$

where $\rho_h^{*(0)} \sim \text{U}(0,1)$, and the acceptance probability α is defined as

$$\alpha\big(\rho_h^{*(d)}, \rho_h^{*(d-1)} \mid \beta_h, r_h, \bar{\rho}^*, \xi_\rho, -\big) = \min\left\{\frac{\pi(\rho_h^{*(d)} \mid \beta_h, r_h, \bar{\rho}^*, \xi_\rho, -)}{\pi(\rho_h^{*(d-1)} \mid \beta_h, r_h, \bar{\rho}^*, \xi_\rho, -)}, 1\right\}.$$

The inverse transformed sample

$$\rho_h^{(d)} = \exp\big(\rho_h^{*(d)}\big)\big/\big(1+\exp(\rho_h^{*(d)})\big)$$

is then employed to satisfy the restriction

$$\big\{S_{kht}(\rho_h^{(d)}) \ge r_h \ge 0, \text{ where } y_{kht} = 1\big\}.$$

4. Hyperparameters:

The draws of the hyperparameters follow standard Bayesian updating.

(4-1) Threshold Parameter:

(1) $\bar{r} \mid \{r_h\}, \nu_r$

$$\bar{r} \sim \text{N}(\delta_r \times \eta_r, \delta_r), \quad \delta_r = (H/\nu_r + 1/\bar{\sigma}_r^2)^{-1}, \quad \eta_r = \nu_r^{-1}\sum_h r_h\Big/H.$$

(2) $\nu_r \mid \{r_h\}, \bar{r}$

$$\nu_r \sim \text{Inverted Gamma (IG)}\ ((\bar{\nu}_r + H)/2, (\bar{\Upsilon}_r + Q_r)/2),$$

$$Q_r = \sum_{h=1}^{H} (r_h - \bar{r})^2.$$

(4-2) Carryover Parameter:

(1) $\bar{\rho}^* \mid \{\rho_h\}, \xi_\rho$

$$\bar{\rho}^* \sim \text{N}(\delta_\rho \times \eta_\rho, \delta_\rho); \quad \delta_\rho = (H/\xi_\rho + 1/\bar{\sigma}_{\rho^*}^2)^{-1}$$

$$\eta_\rho = \xi_\rho^{-1} \sum_h \rho_h / H.$$

(2) $\xi_\rho \mid \{\rho_h\}, \bar{\rho}^*$

$$\xi_\rho \sim \text{Inverted Gamma}\ ((\bar{\xi}_\rho + H)/2, (\bar{\Upsilon}_\rho + Q_\rho)/2),$$

$$Q_\rho = \sum_{h=1}^{H} (\rho_h - \bar{\rho}^*)^2.$$

5. Steps 1–4 above are repeated to obtain the joint posterior density:

$$f\big(\{\beta_h\}, \bar{\beta}, \Lambda, \{r_h\}, \{\rho_h^*\}, (\bar{r}, \nu_r), (\bar{\rho}^*, \xi_\rho) \mid \{\text{PData}_h\}, \{\text{AData}_h\}\big).$$

In the above, we set prior distributions on hyperparameters:

$$\begin{aligned}\pi(\bar{\beta}) &= \text{N}(0, 100I); \quad \pi(\Lambda) = \text{IW}(13, 13I);\\ \pi(\bar{r}) &= \text{N}(0, \bar{\sigma}_{r^*}^2) = \text{N}(0, 100);\\ \pi(\bar{\nu}_r) &= \text{IG}(\bar{\nu}_r/2, \bar{\Upsilon}_r/2) = \text{IG}(6, 1);\\ \pi(\bar{\rho}^*) &= \text{N}(0, \bar{\sigma}_{\rho^*}^2) = \text{N}(0, 100);\\ \pi(\bar{\xi}_\rho) &= \text{IG}(\bar{\xi}_\rho/2, \bar{\Upsilon}_\rho/2) = \text{IG}(6, 1).\end{aligned}$$

References

Allenby, G. M., J. L. Ginter. 1995. The effects of in-store displays and feature advertising on consideration sets. *Internat. J. Res. Marketing* **12**(1) 67–80.

Allenby, G. M., M. J. Garratt, P. E. Rossi. 2010. A model for trade-up and change in considered brands. *Marketing Sci.* **29**(1) 40–56.

Andrews, R. L., T. C. Srinivasan. 1995. Studying consideration effects in empirical choice models using scanner panel data. *J. Marketing Res.* **32**(1) 30–41.

Ansari, A., C. F. Mela. 2003. E-customization. *J. Marketing Res.* **40**(2) 131–145.

Bass, F. M., D. G. Clarke. 1972. Testing distributed lag models of advertising effects. *J. Marketing Res.* **9**(3) 298–308.

Bemmaor, A. C. 1984. Testing alternative econometric models on the existence of advertising threshold effect. *J. Marketing Res.* **21**(3) 298–308.

Bronnenberg, B. J. 1998. Advertising frequency decisions in a discrete Markov process under a budget constraint. *J. Marketing Res.* **35**(3) 399–406.

Bronnenberg, B. J., W. Vanhonacker. 1996. Limited choice sets, local price response, and implied measures of price competition. *J. Marketing Res.* **33**(2) 163–173.

Chakravarti, A., C. Janiszewski, G. Ülkümen. 2006. The neglect of prescreening information. *J. Marketing Res.* **43**(4) 642–653.

Chandukala, S. R., J. P. Dotson, J. D. Brazell, G. M. Allenby. 2009. Bayesian analysis of hierarchical effects. Working paper, Fisher College of Business, Ohio State University, Columbus.

Chessa, A. G., J. M. J. Murre. 2007. A neurocognitive model of advertisement content and brand name recall. *Marketing Sci.* **26**(1) 130–141.

Clarke, D. G. 1976. Econometric measurement of the duration of advertising effect on sales. *J. Marketing Res.* **13**(4) 345–357.

Dubé, J.-P., G. J. Hitsch, P. Manchanda. 2005. An empirical model of advertising dynamics. *Quant. Marketing Econom.* **3**(2) 107–144.

Erdem, T. 1996. A dynamic analysis of market structure based on panel data. *Marketing Sci.* **15**(4) 359–378.

Erdem, T., M. P. Keane. 1996. Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets. *Marketing Sci.* **15**(1) 1–20.

Erdem, T., M. P. Keane, B. Sun. 2008. A dynamic model of brand choice when price and advertising signal product quality. *Marketing Sci.* **27**(6) 1111–1125.

Erickson, G. M. 1997. Note: Dynamic conjectural variations in a Lanchester oligopoly. *Management Sci.* **43**(11) 1603–1608.

Fader, P. S., L. McAlister. 1990. An elimination by aspects model of consumer response to promotion calibrated on UPC scanner data. *J. Marketing Res.* **27**(3) 322–332.

Feinberg, F. M. 2001. On continuous-time optimal advertising under S-shaped response. *Management Sci.* **47**(11) 1476–1487.

Guadagni, P. M., J. D. C. Little. 1983. A logit model of brand choice calibrated on scanner data. *Marketing Sci.* **2**(3) 203–238.

Gilbride, T. J., G. M. Allenby. 2004. A choice model with conjunctive, disjunctive, and compensatory screening rules. *Marketing Sci.* **23**(3) 391–406.

Hanssens, D. M., L. J. Parsons, R. L. Shultz. 2001, *Market Response Models: Econometric and Time Series Analysis*, 2nd ed. Kluwer Academic Press, Boston.

Horsky, D., S. Misra, P. Nelson. 2006. Observed and unobserved preference heterogeneity in brand-choice models. *Marketing Sci.* **25**(4) 322–335.

Howard, J. A., J. N. Sheth. 1969. *The Theory of Buyer Behavior*. John Wiley & Sons, New York.

Jedidi, K., C. F. Mela, S. Gupta. 1999. Managing advertising and promotion for long-run profitability. *Marketing Sci.* **18**(1) 1–22.

Kanetkar, V., C. B. Weinberg, D. L. Weiss. 1992. Price sensitivity and television advertising exposures: Some empirical findings. *Marketing Sci.* **11**(4) 359–371.

Keane, M. P. 1997. Current issues in discrete choice modeling. *Marketing Lett.* **8**(3) 307–322.

Khaneman, D., A. Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* **47**(2) 263–291.

Krugman, H. E. 1972. Why three exposures may be enough. *J. Advertising Res.* **12**(6) 21–25.

Lambin, J. J. 1976. *Advertising, Competition, and Market Conduct in Oligopoly Over Time*. Elsevier, New York.

Lavidge, R. J., G. A. Steiner. 1961. A model for predictive measurements of advertising effectiveness. *J. Marketing* **25**(6) 59–62.

Little, J. D. C. 1979. Aggregate advertising models: The state of the art. *Oper. Res.* **27**(4) 629–667.

Liu, Q., T. Otter, G. M. Allenby. 2007. Investigating endogeneity bias in marketing. *Marketing Sci.* **26**(5) 642–650.

Lodish, L. M., M. Abraham, S. Kalmenson, J. Livelsberger, B. Lubetkin, B. Richardson, M. E. Stevens. 1995. How T.V. advertising works: A meta-analysis of 389 real world split cable T.V. advertising experiments. *J. Marketing Res.* **32**(2) 125–139.

Mehta, N., X. Chen, O. Narasimham. 2008. Informing, transforming, and persuading: Disentagling the multiple effects of advertising on brand choice decisions. *Marketing Sci.* **27**(3) 334–355.

Mehta, N., S. Rajiv, K. Srinivasan. 2003. Price uncertainty and consumer search: A structural model of consideration set formation. *Marketing Sci.* **22**(1) 58–84.

Mela, C. F., S. Gupta, D. R. Lehmann. 1997. The long-term impact of promotion and advertising on consumer brand choice. *J. Marketing Res.* **34**(2) 248–261.

Mizik, N., R. Jacobson. 2008. The financial value impact of perceptual brand attributes. *J. Marketing Res.* **45**(1) 15–32.

Naik, P. A., M. K. Mantrala, A. G. Sawyer. 1998. Planning media schedules in the presence of dynamic advertising quality. *Marketing Sci.* **17**(3) 214–235.

Pedrick, J. H., F. S. Zufryden. 1991. Evaluating the impact of advertising media plans: A model of consumer purchase dynamics using single-source data. *Marketing Sci.* **10**(2) 111–130.

Rossi, P. E., G. M. Allenby, R. McCulloch. 2005. *Bayesian Statistics and Marketing*. John Wiley & Sons, Chichester, UK.

Rossi, P. E., G. M. Allenby, R. McCulloch. 1996. The value of purchase history data in target marketing. *Marketing Sci.* **15**(4) 321–340.

Rubin, D. B. 1976. Inference and missing data. *Biometrika* **63**(3) 581–592.

Simon, J. L., J. Arndt. 1980. The shape of the advertising response function. *J. Advertising Res.* **20**(4) 11–28.

Steiner, R. L. 1987. Point of view: The paradox of increasing returns to advertising. *J. Advertising Res.* **27** 45–53.

Strong, E. K., Jr. 1925. Theories of selling. *J. Appl. Psych.* **9**(1) 75–86.

Tellis, G. J. 2004. *Effective Advertising: Understanding When, How and Why Advertising Works*. Sage, Thousand Oaks, CA.

Vakratsas, D., T. Ambler. 1999. How advertising works: What do we really know? *J. Marketing* **63**(1) 26–43.

Vakratsas, D., F. M. Feinberg, F. M. Bass, G. Kalyanaram. 2004. The shape of advertising response functions revisited: A model of dynamic probabilistic thresholds. *Marketing Sci.* **23**(1) 109–119.

van Nierop, E., B. Bronnenberg, R. Paap, M. Wedel, P. H. Franses. 2010. Retrieving unobserved consideration sets from household panel data. *J. Marketing Res.* **47**(1) 63–74.

Villas-Boas, J. M. 1993. Predicting advertising pulsing policies in an oligopoly: A model and empirical test. *Marketing Sci.* **12**(1) 88–102.