Research Paper No. 1935

The Impact of Feature Advertising on Customer Store Choice

Anand V. Bodapati
V. Srinivasan

May 2006

Research Paper Series

STANFORD
GRADUATE SCHOOL OF BUSINESS

http://ssrn.com/abstract=901458
The Impact of Feature Advertising on Customer Store Choice

Anand V. Bodapati
UCLA Anderson Graduate School of Management

V. Srinivasan
Graduate School of Business
Stanford University

Abstract

A heavily used competitive tactic in the grocery business is the weekly advertising of price reductions in newspaper inserts and store fliers. Store managers commonly believe that advertisements of price reductions and loss leaders help to build store traffic by diverting customers from competing stores, thereby increasing store volume and profitability. It is therefore not surprising that grocery retail planners across competing stores expend considerable thought on what items to advertise each week and at what levels of prominence. What is surprising, however, is that we marketing scientists do not know much about the manner and extent to which feature advertising in a competitive environment influences where and how customers shop. The marketing science literature has not even been able to establish that feature advertising has a substantial impact on store choice, let alone the more operational question of which categories are better at drawing consumers away from one store and into a competing store. In this paper we employ a stochastic choice modeling framework to propose and empirically estimate a disaggregate, consumer-level model of the effects of feature advertising on store choice. We use this model to understand which categories are more influential drivers of store traffic and better at diverting consumers from competing stores.
1 Introduction

A heavily used competitive tactic in the grocery business is the weekly advertising of price reductions in newspaper inserts and store fliers, hereafter referred to as feature advertising. Store managers commonly believe that advertisements of price reductions and loss leaders help to build store traffic by diverting customers from competing stores, thereby increasing store volume and profitability. Our conversations with grocery managers suggest that they value sales increases that are the result of increased store traffic more than they value sales increases that are not the result of increased traffic. This is because the former are seen as displacement of sales from competitors and hence representing a “true” sales increase, whereas the latter as seen primarily as displacement of sales from the store’s own future-period sales (due to purchase acceleration, for instance) or same-period sales (due to brand-switching, say). There are many tactics used by grocery stores to increase store traffic: feature advertising, television advertising, and targeted direct mail communications are the three most common. Of these, feature advertising is perceived to be the most cost-effective way to deliver information that would influence consumers’ store choices each week. So strong is this belief that in 2002, U.S. supermarkets spent a staggering $8 billion on feature advertising; this figure represents approximately 2% of 2002 sales and is very large, considering that net profit before taxes was only 2.23% of sales. Further, feature advertising accounts for about one half of the total advertising spending in the supermarket business. Television advertising, which is the next largest component, accounts for only about 8% of advertising spending in this industry.¹ It must be mentioned that a considerable fraction of this $8 billion comes from manufacturers’ subsidies; nevertheless, the use of manufacturers’ subsidies is to a large extent discretionary in that retailers can typically use the subsidies in ways other than on feature advertising. Therefore, that retailers choose to disburse these funds towards feature advertising is suggestive of retailers considering feature advertising to be an effective vehicle for increasing store profitability.

Given the situation presented above, it is not surprising that grocery retail planners across competing stores expend considerable thought on what items to advertise each week and at what levels of prominence. Nor is it surprising that managers in the grocery industry seek a better understanding of the impact of feature advertising on sales. This leads to the following research question for marketing scientists: In a competitive setting, how does a store’s feature advertising influence the store choices of customers in its trading area? Our paper addresses this important issue, to which the marketing literature has paid scant attention. Indeed, the attention has been

¹Sources: The Food Marketing Institute’s publications, Key Facts May 2002, and Supermarket Research, May-June 2002
so rare that there is no research yet that demonstrates that feature advertising has any appreciable influence on store traffic in the grocery business, let alone the more operational question of which categories are better at drawing consumers away from one store and in to a competing store.

The attention has been scant, we believe, because measuring the effect of feature advertising is really a difficult task. The task is difficult because of three factors, we would argue: (a) The fraction of consumers whose store choices are regularly driven by feature advertising is small. (b) Even amongst consumers who do regularly use feature advertising in the store choice decision, there is very likely high heterogeneity in which categories they pay attention to. For instance, some consumers may look at feature advertisements in the soft drinks category to decide on which store to go to whereas some other customers may look for advertisements in the paper towels and frozen pizza categories. The result of this is that feature advertising in any one category may have only a very weak effect on the average consumer’s store choice decision in any one week. Finally, (c) a feature advertisement is almost always accompanied by a price reduction which creates a large effect on sales on its own. Factors (a), (b) and (c) operating together create a problem of very low signal to noise ratio in that the fraction of across-time variance in sales attributable only to the traffic building effect of feature advertising can be rather small and undetectable unless the statistical modeling is done carefully. In this paper, we attempt a careful statistical modeling which we believe goes further than past research in helping us uncover and understand the effects of feature advertising on store choice.

This research employs a stochastic choice modeling framework to propose and empirically estimate a disaggregate, consumer-level model of the effects of feature advertising and accompanying promotions on a store’s sales, in the face of competitive activities from other stores. Our stochastic choice model, when estimated and parameterized on the basis of real data, potentially offers us several insights on the effectiveness of retailers’ competitive actions in feature advertising. This model permits us to investigate the extent to which retailer promotion programs divert customers from competing stores. Consequently, it offers insights into how the retailer may deploy price promotions and feature advertising more effectively. The structural properties of the estimated model allow us to give some normative characterizations on which categories ought to be the most actively feature-advertised.

There is only limited past research examining the effect of feature advertising on store choice. We now review the small literature that does exist and relate the existing papers to the present paper. Kumar and Leone (1988) consider the hypothesis that a brand’s promotions (price reductions, feature advertising and special display) increase sales of the brand in the store, and decrease the sales of
competing brands in the store and of both competing- and same-brand sales in competing stores. Using aggregate level sales data for three brands in the children’s disposable diapers category, they find limited support for these hypotheses. Their results show that there exist some cases where promotions for a brand decrease sales in competing stores and increase the brand’s sales in the store which does the advertising. However, the Kumar and Leone (1988) analysis leaves unexplained why the sales displacement is happening. Is it because promotions (including feature advertising) are causing people to switch stores and displace sales? Or is it because people are choosing (switching) stores in a promotions-independent manner but buying items (once within the store) opportunistically in a promotions-dependent manner and hence displacing sales from non-promoting stores to promoting stores? Walters (1991) examines the same hypotheses but extends to the effects on sales of complementary products and categories. Using a different model, also on aggregate level sales data, he considers sales of brands in four categories: cake mix, cake frosting, spaghetti, and spaghetti sauce. He considers only price promotions and not feature advertising or display. He notes, however, that price reductions are highly correlated with these forms of advertising and implies that the effects found of price promotion ought to be true for advertising as well. Walters’s results are similar to those of Kumar and Leone. There exist some instances (though this does not hold for most instances) where price reductions of a brand in one store cause sales reductions of same-brand and competing-brands in other stores. This generalizes the Kumar and Leone results in that Walters shows that price reductions in a store in a certain category can cause reductions in competing-store sales even outside that category. Even if one assumes that Walters’s results would hold true for feature advertising as well as price reductions (recall that he considers only price reductions), one is left with the same ambiguity posed by the Kumar and Leone results. Is the store displacement happening because of consumers switching stores in response to out-of-store promotions? Or are they not responding in store choice to out-of-store promotions but only to in-store promotions once they are in the store?

It is important to distinguish between these two possibilities because depending on which scenario is more prevalent, one of these opposing strategies would be a better one: (i) if sales displacement is being driven largely by store-switching as affected by out-of-store advertising, then the store manager ought to shift emphasis onto feature advertising as a promotional vehicle, (ii) if out-of-store advertising has only a weak effect on store choice, and if there is a large number of consumers shopping across multiple stores independent of feature advertising, then, the store manager ought to increase the level of in-store promotions to maximize opportunistic buying.

Walters and MacKenzie (1988) consider deep-discounted, feature-advertised items in eight
categories. They use a structural equations model that explicitly distinguishes store traffic and store sales. This model therefore is capable of disambiguating scenarios (i) and (ii) identified above. Their results however show that the feature-advertised loss-leaders do not increase store traffic at all (the one exception in the eight categories is “rolls and buns”, where the coefficient is just significant at the 5% level). This is a surprising result that has weak face validity because the eight categories they consider are those that are generally deep-discounted by stores and believed to be effective traffic-builder categories. We submit the argument that this work was unable to detect significant effects of feature advertising on store traffic because it used aggregate level store data where the “low signal to noise ratio” problem we identified earlier is even higher than with panel data. With aggregate data, the only way to study dependencies is via across-time sales variation; However, most this this variation probably gets attributed to the price-promotions accompanying the feature advertising, with very little marginal effect attributable to sales from increased store traffic. Unlike the Walters and MacKenzie work, the present research uses panel data which affords study of both across-time and across-consumer variation and therefore offers a better opportunity to tease out the effects of feature advertising.

There are three important papers which, like the present paper, look at store choice at the individual consumer level: Bell & Lattin (1998), Bell, Ho & Tang (1998) and Rhee & Bell (2002). The paper by Bell & Lattin (1998) is an important work with substantial implications for supermarket managers. The paper argues that consumers whose trips tend to be infrequent and large will see higher expected basket attractiveness in EDLP stores than in HILO stores. To test this hypothesis, Bell & Lattin estimate store choice as a function of expected basket attractiveness using a mixture model over two classes of customers, one class for large-basket shoppers and another class for small-basket shoppers. The model controls for several variables, both time-variant and time-invariant. Feature advertising is captured via a single, cleverly-constructed variable consisting of a weighted average of feature advertising levels for all brands in all categories at the time of the store choice. This predictor is introduced into the store choice model largely as a control variable rather than as a construct of focal interest. For this reason, Bell & Lattin do not consider issues as we do such as the across-category differences in the impact of feature advertising on store choice.

Bell, Ho & Tang (1998) model the utility of a store to an individual consumer with a time-invariant fixed-cost component and a time-varying variable-cost component. The variable cost is taken as the sum over all items of the expected cost of that item multiplied by a 0-1 binary variable taking value 1 if the item is on the consumer’s shopping list, which is seen to be composed prior to the store choice decision and therefore prior to the store visit. The consumer’s decision process in
this paper does not consider time-specific prices to be decision factors, conditional on the shopping list. However, because the value of the above binary variable is unknown to the analyst (though assumed known to the consumer), it is imputed from the store price environment and from what the consumer is known to have bought on the trip; because the store prices inform the imputation of the binary variables, the store choice model unconditional on the shopping list ends up being a function of prices. It is this unconditional model that is estimated by Bell, Ho & Tang and so the final store choice model is a function of the stores’ prices. Their paper is very interesting because the estimated model lends itself to other interpretations and many effects are nicely represented in the model. However, feature advertising does not explicitly enter into the store choice model. (We will say more in the next paragraph about our use of the word “explicitly” in this last sentence.) Rhee & Bell (2002) is an interesting paper that studies the “mobility” of consumer store choice: the paper models the probability that the store most patronized by a consumer in one week is different from the store most patronized in the previous week, as a function of time-invariant household-specific variables and time-variant factors. One time-variant predictor that is considered is the price of a certain basket of items (the same basket is considered for all weeks, all stores and all households), which is found to be statistically insignificant. As in the Bell, Ho and Tang (1998) paper, this model does not explicitly consider the effect of feature advertising on store choice.

While the models in Bell, Ho & Tang (2002) and Rhee & Bell (2002) do not explicitly incorporate feature-advertising in predicting store choice, they do incorporate variables that are direct functions of items’ prices that the consumer would encounter once she entered the store. To the extent that these in-store prices of items are supersets of and correlated with the respective out-of-store advertising levels, the models in these two papers can be seen to capture the effects of feature advertising on store choice. However, it is important to recognize the conceptual distinction between in-store prices and out-of-store advertising levels, particularly for modeling the store choice decision and the behavioral story that goes with it. It is important to recognize that the store choice decision is being made prior to the consumer entering the store and therefore the consumer will typically not know the prices of all items in the store at the time of the decision. At the time of the store choice decision, the consumer will typically have access to price information on only those items whose prices have been revealed through feature advertising and, based on this information, she will need to form expectations about the prices that she would encounter in the store were she to visit the store. Most papers on store choice in the literature, use all items prices (feature-advertised or not) as predictors in the store choice model. An important contribution of our paper is in the behavioral story we put forward for how feature advertising affects price expectations, not just for the advertised items...
but also of the unadvertised items, and how these price expectations influence store choice. This conceptualization allows us to propose a model where store choice is only a function of feature advertised prices and not a function of items’ prices that the consumer would usually not have access to at the time of the store choice decision. The conceptualization is followed through to estimate a model which allows different categories’ feature advertisements to have different impacts on store choice, thereby helping the manager assess the differential efficacies of feature advertising in various categories.

The rest of this paper is written as follows. Section 2 develops the core model for store choice in the framework of random-utility models. Section 3 presents the chief empirical results. We close the paper in Section 4 with a prospective view of future work and how this paper could contribute to our understanding of other problems in retailer promotions.

2 A Model for the Influence of Feature Advertising on Store Choice

Our model attempts to tackle head-on the “low signal-to-noise ratio” problem we discussed earlier, which we believe has made it difficult for past research efforts to isolate the effect of feature advertising on store choices. We submit that the “low signal-to-noise ratio” is created because there are only some situations where feature advertising impacts store choice and that in other situations, which are probably far more common, we would not expect feature advertising to influence store choices. Therefore, if we were to pool observations across all situations, we would find it difficult to isolate the influence of feature advertising on store choices. The isolation would be much easier if we could purify the set of observations and concentrate on just those observations where feature advertising is likely to have an effect. This is exactly what we attempt to do. Doing this, however, is not easy because we do not directly observe and know that a certain situation did contain an influence of feature advertising on store choice; rather we need to probabilistically impute the occurrence of such an event based on empirical prior beliefs and observed events. This is what complicates the modeling and the estimation and therefore necessitates careful statistical analysis.

To allow the model to focus attention on situations where feature advertising will potentially influence store choice, we incorporate into it three important ideas:

(A) Consumers vary in the strategies they employ to reduce grocery expenditure. Only some consumers use feature advertising to decide which store minimizes the cost of the anticipated expenditure.
purchases on a given trip. The Food Marketing Institute’s survey-based report on consumer
trends in 2000 says that 24% of the consumers say they “never” look in the newspaper for
grocery specials, that 23% say they do it “occasionally” and that 54% say they do it “fairly
often” or “pretty much every time”. The same survey research reports that while 73% of
consumers say they only occasionally or never switch supermarkets in response advertised
specials, 28% say they do this “pretty much every time” or “fairly often”. These numbers
suggests that there is considerable heterogeneity in whether consumers use feature advertising
to influence their store choices.

(B) The extent to which a certain shopping trip is a major trip or fill-in trip may influence which store
is chosen on that trip. For instance, a consumer may ordinarily go to a nearby convenience
store if he or she wants only a carton of milk, but may prefer a supermarket for the weekly
stock-up trip. It is easy to see why this phenomenon would hold. We can view the cost of
acquiring a basket of goods at a store as the sum of (1) the visiting cost of the store, and (2)
the basket cost. Bell and Lattin (1998) use a similar decomposition for costs. We use the term
“visiting cost” to refer to all the fixed costs associated with visiting the store. Visiting cost
will depend on factors like store accessibility and the consumer’s core store preferences. The
basket cost, the price paid for the basket in the store is a variable cost in that it depends on the
volume and content of the basket. On an extreme fill-in trip, the volume of purchases is very
small. Hence, the acquisition cost is dominated by the visiting cost. On the other hand, for
a major shopping trip, the basket cost dominates the acquisition cost. The key point is this:
If the ordering of stores according to visiting cost is different from the ordering according to
expected basket cost, then we would expect that the ordering of stores according to the total
basket acquisition cost will be a function of the expected basket cost.
Among consumers who sometimes use feature advertising for their store choice decisions, feature advertising is more likely to influence store choice decisions on major trips than on fill-in trips. The argument for this closely follows the logic given for idea (B). Total basket acquisition cost is a function of visiting cost and basket cost. Any price discounts the customer is made aware of via a feature advertisement would influence only his/her estimate of the basket cost. On larger trips, the basket cost would have a greater influence on the store choice decision. Therefore, we would expect that feature advertising would have more of an influence on the store choice decision on major trips.

The net result of Ideas (A), (B) and (C) identified above is that we would expect feature advertising to impact the store choice decision for only some consumers and that too for only some kinds of trips, and it is only on such trips that the model should attempt to fit an influence of advertising on store choice. Attempting to fit an influence of advertising on store choice for all customers for all trips would dilute the effect of advertising on store choice and make it far more difficult to detect. We now discuss the incorporation of ideas (A), (B) and (C) into a stochastic choice model.

2.1 Modeling Idea (B): The trip’s anticipated purchase volume may influence store choice

To incorporate Idea B, we need a measure that captures the extent to which a shopping trip is anticipated by the consumer to be a major trip versus a fill-in trip. The measure we propose to use is $I^h_t$, an estimate of the household’s total inventory of grocery goods just before trip $t$, measured in dollar value. $I^h_t$ is computed through the relation

$$I^h_t = \eta^h I^h_{t-1} + (V^h_{t-1} - L^h_t R^h), \quad 0 \leq \eta^h \leq 1.$$  \hfill (1)

The change in inventory between two successive trips is $V^h_{t-1}$, the dollar value of the basket bought on the earlier trip, less the consumption $L^h_t R^h$ in the intervening time. $L^h_t$ is the number of days between the two trips, and $R^h$ is the daily consumption rate of the household. The factor $\eta$ is used to weigh recent flow more heavily and to reduce sensitivity to initial conditions and trips taken much earlier. The basic idea behind estimating the household’s total inventory is simple: the larger and the more recent the previous trips, the lower the likelihood that the current trip is a major shopping trip. The consumption rate $R^h$ can be estimated through a hold-out sample. The value of the weight factor $\eta^h$ is to be chosen to minimize the variance of the estimates (across all shopping trips) of the inventory immediately after a trip. This choice is motivated by the idea that on each trip a consumer
buys a volume that is just sufficient to bring total inventory (post-trip) to an approximately constant “reservoir level”.

We posit that a consumer’s store selection follows a stochastic process, in which the choice on one occasion is independent of the choice on another occasion, but dependent on the extent to which the trip is a major trip or fill-in trip. We write the probability \( \rho_{ts}^h \) that household \( h \) visits store \( s \) on trip \( t \) as

\[
\rho_{ts}^h = (1 - M_t^h) \phi_s^h + M_t^h \mu_s^h(t),
\]

where

\[
M_t^h = \frac{1}{1 + \exp(\psi_0^h + \psi_1^h I_t^h)}, \quad 0 \leq \psi_1^h.
\]

\( M_t^h \) is the degree to which the trip is a major shopping trip versus a fill-in trip. Note that \( M_t^h \) varies from 0 to 1 and is a decreasing function of \( I_t^h \). The \( \psi_0^h \) and \( \psi_1^h \) are household specific parameters whose values are estimated from the data. We can see that the probability \( \rho_{ts}^h \) varies between \( \phi_s^h \) and \( \mu_s^h(t) \) and according to the value of \( M_t^h \). The \( \mu_s^h \) and \( \phi_s^h \) may be interpreted as the probabilities of choosing store \( s \) on, respectively, an extreme major trip and an extreme fill-in trip.

Note that \( M_t^h \) and its key argument \( I_t^h \) attempt to assess the likelihood of the trip being a major shopping trip using only information available prior to the trip and not information on the actual volume of purchases on the trip. That is because, to predict store choice, the construct we use is the extent to which the trip is anticipated to be a major trip rather than what the trip actually was. The store choice decision is made by the consumer before the trip is made and so we use only information available prior to commencement of the trip.

### 2.2 Modeling Ideas (A) and (C): use of feature-advertising in store choice by only some consumers and more likely on major trips

To allow Idea A, we posit that there are two kinds of consumers: (1) Non-Responders, whose store choices are not affected by feature advertising, and (2) Responders, whose store choices are affected by feature advertising. It is important to note that a Non-Responder is not necessarily a loyal customer, he may well be a “Non-Loyal” (see the typology in Figure 1). For both types of shoppers, the probabilities of store choice on an extreme fill-in trip (i.e., the values of \( \{\phi_s^h\} \)) are invariant with time and, in particular, invariant with feature advertising by competing supermarkets. This aspect of the model is motivated by the assumption that the basket value of the extreme fill-in trip is typically very low. Therefore, any savings from possible price promotions is small and, in
particular, assumed to be too small to cover the information-processing cost of inspecting competing stores’ feature advertising to decide where to shop on that trip. We model the probabilities \( \{ \phi^h_s \} \) using a logit formula as

\[
\phi^h_s = \frac{\exp(\alpha^h_{\phi_s})}{\sum_{s' \in S} \exp(\alpha^h_{\phi_{s'}})},
\]

where the \( \alpha^h_{\phi_s} \) are household-level parameters to be estimated from the data. \( S \) is the set of stores in the market. The difference between the two types of shoppers arises in their decision rules for larger baskets of planned purchases: For Responders, the store choice probabilities \( \{ \mu^h_s(t) \}_{s \in S} \) on an extreme major trip are a function of an occasion-specific measure of store attractiveness which depends on the advertised prices at that time. For Non-Responders, the \( \{ \mu^h_s(t) \}_{s \in S} \), like the \( \{ \phi^h_s \}_{s \in S} \), are time-invariant.\(^2\) Let \( z(h) \) be a 0-1 variable taking value 0 if household \( h \) is a Non-Responder.

\(^2\)A consumer may be thought of as a Responder or a Non-Responder according to his/her information-processing cost of doing between-store price comparison — before each major trip — based on prices in feature advertising. If the cost is less than the expected savings from price-comparison shopping, the consumer becomes a Responder. Otherwise, the consumer becomes a Non-Responder.
and the value 1 for a Responder. We write the store choice probabilities in an extreme major trip as

\[
\mu^h_s(t) = \begin{cases} 
\frac{\exp(\alpha^h_{\mu s})}{\sum_{s' \in S} \exp(\alpha^h_{\mu s'} + S^h_{s'}(t))} & \text{if } z(h) = 0, \text{ Non-Responder} \\
\frac{\exp(\alpha^h_{\mu s} + S^h_{s}(t))}{\sum_{s' \in S} \exp(\alpha^h_{\mu s'} + S^h_{s'}(t))} & \text{if } z(h) = 1, \text{ Responder.} 
\end{cases}
\]

\(S^h_{s}(t)\) is a measure of the attractiveness of store \(s\) to the household at time \(t\).

We model the store attractiveness for Responders as

\[
S^h_{s}(t) = \sum_{c \in C} \alpha^h_{\mu c} EECA^h_{sc}(t),
\]

\(EECA^h_{sc}(t) = E_{\text{prices}_{tc}}(t) \left[ CA^h_{tc} \times p^h_{i|s} | \text{prices}_{sc}(t) \right].
\]

The two equations above contain many elements which we now explain. The store attractiveness \(S^h_{s}(t)\) is the weighted sum, over all categories, of a measure of the attractiveness of a store’s offerings in each category. We refer to the measure as the externally evaluated category attractiveness (EECA). The \(c\) denotes a category and \(C\) denotes the set of all categories. The weight \(\alpha^h_{\mu c}\) may be interpreted as the influence of promotions in category \(c\) (relative to the influence of promotions in other categories) on where household \(h\) shops on a given shopping trip\(^3\); for example, a consumer who chooses stores according to promotions in the pizza and soft drinks categories (rather than other categories) will have high values of \(\alpha^h_{\mu \text{pizza}}\) and \(\alpha^h_{\mu \text{softdrinks}}\). Note that the model of \(\mu^h_s(t)\) for Non-Responders is a special case of the model for Responders, with \(S^h_{s}(t)\) being the same across all stores.

Our measure of the externally evaluated category attractiveness is almost the same as \(CA^h_{tc}\), the inclusive value term used in the usual nested logit type model of purchase incidence. (See Guadagni and Little (1983), Bucklin and Lattin (1987). See section 2.3.3.) There is one crucial difference. The inclusive value \(CA^h_{tc}\) is a function of all prices and promotions that constitute the store environment. The consumer, however, does not have full knowledge of the store environment at the time of the store choice decision, because this information is available only if the consumer visits the store. Therefore, the \(CA^h_{tc}\) cannot directly be used as a measure of the externally evaluated category attractiveness. (The qualification “externally evaluated” emphasizes that the category attractiveness is evaluated before entering the store.) However, the consumer does have partial knowledge about the store environment, based on past experience and on price information from feature advertisements. This partial knowledge can be used to inform a probability distribution

\(^3\)Strictly speaking, the \(\{\alpha^h_{\mu c}\}_{c \in C}\) also serve to scale the attractiveness values for all categories to comparable units. Therefore, a quantity like \(\alpha^h_{\mu} \times \text{std.dev.}(EECA^h_{sc}(t))\) is a more correct measure of category importance.
for the values defining the store environment. Our measure for the externally evaluated category attractiveness is simply the expectation of $CA_{tc}^h$ over this probability distribution:

$$E_{\text{prices}_{sc}(t)}[CA_{tc}^h \times p_t^h(i_c|s)\mid \text{prices}_{sc}(t)].$$

The term inside the expectation operator is the expected utility from the offerings of category $c$ in store $s$. It can be shown (following Ben-Akiva and Lerman 1985) that $CA_{tc}^h$ is the expected maximum utility under a brand-choice model with Weibull random utilities. Therefore, assuming that the consumer chooses the item in $c$ with the greatest utility, $CA_{tc}^h$ is the expected utility from $c$ conditional on a purchase. The term $p_t^h(i_c|s)$ is the category incidence probability, the likelihood that the consumer makes a purchase in $c$. Standardize the utility from making no purchase to be 0. Therefore the expected utility from category $c$ (unconditional on any purchase) is

$$CU_{tc}^h = CA_{tc}^h \times p_t^h(i_c|s) + 0 \times (1 - p_t^h(i_c|s))$$

$$= CA_{tc}^h \times p_t^h(i_c|s).$$

The mathematical expressions for the purchase incidence probability $p_t^h(i_c|s)$ and category attractiveness $CA_{tc}^h$ as a function of prices are described in section 2.3.3.

### 2.3 Estimating the Externally Evaluated Category Attractiveness

The above expectation over the probability distribution of the vector of prices defining the store environment is estimated nonparametrically using the simple but powerful technique of importance reweighting. We now describe this method first in symbolic algebra and then via a numerical example in section 2.3.1. Let $P_{sc}$ denote the random row-vector containing the price information for all items in category $c$ in store $s$. Let the dimensionality of $P_{sc}$ be $d$. Given observed vectors $\{P_{sc}^n\}_{n=1}^N$ for $N$ weeks, the nonparametric Gaussian product-kernel estimate for the density of $P_{sc}$ is (Scott 1992, page 152 ff):

$$f(P_{sc}) = \sum_{n=1}^{N} \frac{1}{N\sqrt{2d\pi d}|H_N|} \exp \left( -\frac{1}{2}(P_{sc} - P_{sc}^n)^T H_N^{-1} (P_{sc} - P_{sc}^n) \right),$$

where $H_N$ is a suitably chosen diagonal bandwidth matrix. Notice that the Gaussian kernel estimate expresses the density simply as a mixture of Gaussian components, each centered on an observed vector. Stated otherwise: we express the density as just the empirical density with random Gaussian perturbations around the empirically observed values of $P_{sc}$. The implication of $H_N$ being diagonal is that these perturbations are independent across brands. The perturbations being independent does
not at all imply that the components of \( \mathbf{P}_{sc} \) are independent. The components will be correlated in \( f(\mathbf{P}_{sc}) \) if they are correlated in the empirically observed values.

Now consider the estimate of the density of \( \mathbf{P}_{sc} \) given values of certain of its components. (This corresponds to the consumer’s expectations about unobserved in-store prices given information on some of the prices through feature advertising.) Assume that the vector \( \mathbf{P}_{sc} \) is written so that

\[
\mathbf{P}_{sc} = [\mathbf{P}_{sc}^{(k)} \mathbf{P}_{sc}^{(u)}],
\]

where \( \mathbf{P}_{sc}^{(k)} \) is the vector of prices known from the store’s feature advertising that week, and \( \mathbf{P}_{sc}^{(u)} \) is the vector of unknown prices. Assume that bandwidth matrix can correspondingly be written as

\[
H_N = \text{diag}[H^{(k)}, H^{(u)}].
\]

It immediately follows from (9) above that, given values \( \mathbf{p}_{sc}^{(k)} \) for the known prices, the conditional density is

\[
f(\mathbf{P}_{sc} | \mathbf{p}_{sc}^{(k)} = \mathbf{p}_{sc}^{(k)}) = \frac{\prod_{n=1}^{N} \frac{1}{\sqrt{2\pi |H_N| }} \exp\left(-\frac{1}{2}((\mathbf{p}_{sc}^{(k)} - \mathbf{P}_{sc}^{(k)})H_N^{-1}((\mathbf{p}_{sc}^{(k)} - \mathbf{P}_{sc}^{(k)})^T)\right)}{\prod_{n=1}^{N} \frac{1}{\sqrt{2\pi |H_N| }} \exp\left(-\frac{1}{2}((\mathbf{p}_{sc}^{(k)} - \mathbf{P}_{sc}^{(k)})H_N^{-1}((\mathbf{p}_{sc}^{(k)} - \mathbf{P}_{sc}^{(k)})^T)\right)} \right) d\mathbf{P}_{sc}^{(u)}
\]

The expectation of any function \( g(\mathbf{P}_{sc}) \) over the above density is shown in Appendix A to be:

\[
E(g(\mathbf{P}_{sc}) | \mathbf{P}_{sc}^{(k)} = \mathbf{p}_{sc}^{(k)}) = \frac{\sum_{n=1}^{N} w_n g(\mathbf{p}_{sc}^{(k)} \mathbf{P}_{sc}^{(u)} n)}{\sum_{n=1}^{N} w_n},
\]

\[
w_n = \exp\left(-\frac{1}{2}((\mathbf{p}_{sc}^{(k)} - \mathbf{P}_{sc}^{(k)} n)^T)H_k^{-1}((\mathbf{p}_{sc}^{(k)} - \mathbf{P}_{sc}^{(k)} n)^T)\right).
\]
<table>
<thead>
<tr>
<th>week</th>
<th>Coke price</th>
<th>Pepsi price</th>
<th>PC Cola price</th>
<th>CA value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7.99</td>
<td>7.99</td>
<td>4.99</td>
<td>-5.744</td>
</tr>
<tr>
<td>2</td>
<td>7.99</td>
<td>5.99</td>
<td>4.99</td>
<td>-2.972</td>
</tr>
<tr>
<td>3</td>
<td>7.99</td>
<td>5.99</td>
<td>4.99</td>
<td>-2.972</td>
</tr>
<tr>
<td>4</td>
<td>7.99</td>
<td>7.99</td>
<td>4.99</td>
<td>-5.744</td>
</tr>
<tr>
<td>5</td>
<td>5.99</td>
<td>7.99</td>
<td>4.99</td>
<td>-5.186</td>
</tr>
<tr>
<td>6</td>
<td>7.99</td>
<td>5.99</td>
<td>4.59</td>
<td>-2.963</td>
</tr>
<tr>
<td>7</td>
<td>5.99</td>
<td>7.99</td>
<td>4.99</td>
<td>-5.186</td>
</tr>
<tr>
<td>8</td>
<td>4.99</td>
<td>7.99</td>
<td>4.59</td>
<td>-4.212</td>
</tr>
<tr>
<td>9</td>
<td>7.99</td>
<td>4.99</td>
<td>4.59</td>
<td>-1.480</td>
</tr>
<tr>
<td>10</td>
<td>5.99</td>
<td>7.99</td>
<td>4.59</td>
<td>-5.107</td>
</tr>
</tbody>
</table>

Table 1: Illustrating the EECA calculation

drinks category. Suppose that the store never promotes both Coke and Pepsi during the same week. Therefore, as a rational consumer, she must conclude that Pepsi is not on promotion that week and have a correspondingly lower value of EECA. This effect is well captured by our method of importance reweighting: The price vectors that do not have Pepsi on promotion will be given larger weights in her EECA computation (because those are the vectors which have Coke on promotion and therefore are closer to what is known in the current situation through the feature advertisement). This effect is illustrated in the numerical example below computing the EECA value.

2.3.1 A numerical example illustrating the EECA calculation

To illustrate the importance sampling approach to compute the EECA, let us consider a customer who has seen, in ten weeks of shopping at a store, prices in the cola drink category as shown in table 1. Prices are reported for a 24-can pack for the three brands in the category: Coke, Pepsi, and President’s Choice (PC) Cola.

Let us suppose that the consumer’s utility for brand \( b \) at price \( p \) is

\[ u_b = \alpha_b - 1.5 \times \text{price} \]

and that \( \alpha_{\text{Coke}} = 3, \alpha_{\text{Pepsi}} = 6, \alpha_{\text{PCCola}} = 0 \). Therefore the category attractiveness \( CA \) value for
the cola drink category in the \( n \)th week is, following the expression for \( CA \) given in section 2.3.3,

\[
CA_n = \log(\sum_b \exp u_b)
\]

Therefore, the \( CA \) value for the first week will be

\[
CA_1 = \log(\exp(3 - 1.5 \times 7.99) + \exp(6 - 1.5 \times 7.99) + \\
+ \exp(0 - 1.5 \times 4.99))
\]

\[
= -5.744.
\]

The last column in table 1 reports the \( CA \) value for each of the ten weeks.

Consider now the consumer’s expectation of prices for the three brands in the store for the eleventh week. Let us say she assumes that there is no temporal structure across weeks in the prices. (This is actually a good assumption. Serial correlations in price series is very weak in scanner datasets released by Nielsen Marketing and Information Resources Inc. If there do exist significant correlations, then the importance sampling scheme can be modified so that the weights reflect temporal structure such as price cyclicity.) Under the assumption of temporal independence, the consumer’s best Bayesian belief distribution for the prices in the eleventh week would simply be the empirical distribution of the ten previously observed price vectors; that is, given no other information, her Bayesian belief distribution would place probability mass \( \frac{1}{10} \) on each of the 10 price vectors. Therefore, the expected value of the category attractiveness, \( CA \) evaluated on this distribution in simply the average of the values we see in the last column of Table 1. The average of these is \(-4.156\). Hence, \(-4.156\) is the category attractiveness that the consumer should expect — given no specific information about prices in the eleventh week — for the cola drink category if she were to visit the store.

Suppose now that the consumer sees a feature advertisement in the eleventh week announcing that the price of Coke in that week is $5.99. The importance sampling idea prescribes that the Bayesian belief distribution for the prices (other than those advertised) conditional on this information be again the empirical distribution but with the probability masses adjusted in favor of price vectors that are more consistent with the feature advertised prices. The adjustment is given by the Gaussian kernel weights of equation 12 as we illustrate now. In Table 2, we show the Bayesian belief distribution of prices after learning about the feature-advertised price. The prescribed Bayesian belief distribution has probability masses at the ten price-vectors shown in the columns 2–4 of Table 2. The first element in each price-vector is simply the announced price of Coke. For the items
Table 2: Illustrating the EECA calculation: the price-vectors in the belief distribution conditional on the featured advertised price (contd).

whose prices are unadvertised we simply consider the historical price-vectors. For price-vectors (over unadvertised and advertised items) so constructed, we can compute the category attractiveness values as before. The fifth column in Table 2 gives the category attractiveness values at each of the ten price-vectors. We need to weigh heavier the historical price vectors that are consistent with the announced prices; the weighting is given by the Gaussian kernel weights of equation 12.

Following a rule of thumb in the nonparametric density estimation literature, we take the Gaussian kernel variance (the $H_k$ of equation 12) to be simply the observed price variance of the conditioning price vector (which in this case is the price of Coke). If the prices of more than one brand within the category are advertised, then we would take $H_k$ to be the diagonal matrix with the diagonal elements being the variances of the the prices of the advertised items. The observed variance of the price of Coke in the ten weeks is 1.433. Therefore, we set the Gaussian kernel variance as $H_k = 1.433$. Following the numerical form of the numerator in equation 12, the unnormalized weight given to the first price-vector is:

$$w'_n = \exp\left(-\frac{1}{2}(7.99 - 5.99)(1.433)^{-1}(7.99 - 5.99)\right)$$

$$= 0.248.$$
The unnormalized weights for the other price-vectors are as reported in the sixth column of Table 2. Normalizing the weights in the sixth column so that they sum to 1, we get the weights defined in equation 12. The last column of Table 2 gives these weights, which represent the probability masses assigned to each of the price vectors in the consumer’s Bayesian belief distribution of prices, given the information in the feature advertising. Consequently, the expected value of the category attractiveness, $CA$ evaluated on this conditional distribution is simply the weighted average of the values in the fifth column of Table 2 where the weights are given by the numbers in the last column. This weighted average turns out to be $-4.659$. Hence, $-4.659$ is the category attractiveness that the consumer should expect — given the announced price of $5.99$ for Coke – for the cola drink category if she were to visit the store. Note that this figure is actually lower than $-4.156$, the number we obtained earlier for the expected category attractiveness given no specific price information for the eleventh week. Why? As we discussed earlier, the correlational structure in the price-vectors suggests that because Coke is on promotion, Pepsi will not be on promotion. Because the consumer has higher affinity for Pepsi, this added information diminished the expected category attractiveness for her.

2.3.2 Further Comments on EECA

Notice that our formulation of the EECA considers the expectation of category utility over the distribution of prices, rather than the category utility at the expected value of the price vector. To see why the second choice is inappropriate consider a store’s soft drinks category where Coke and Pepsi are both “regular” priced at $8.00 and promoted every week, but never both together, at $6. For a non-loyal consumer, the effective category price at every existing price vector is $6. However, the effective category price for her at the mean price vector would be $7, which would lead to an incorrect evaluation of category attractiveness. This problem arises of course because the second choice ignores the correlational structure of prices. (Note that the method of importance reweighting does the right thing here.)

Strictly speaking, the EECA is meant to represent the expectation of category attractiveness over the distribution representing the consumer’s price knowledge, given feature advertising. However, our computation above sets EECA to the expectation over the actual conditional price distribution (or at least, the statistician’s best estimate of it). Are we assuming that the consumer’s price knowledge coincides with actuality in all stores for all items in all categories? In computational terms, Yes, but in practical terms, No. The algebraic behavior of the $CA_{hc}^b$ term is such that it is fairly insensitive
to price variations in items for which the consumer has low utility (and hence buys infrequently)\(^4\). Therefore, the accuracy of the price distribution imputed for infrequently purchased items does not affect the EECA much. This means that the broad validity of the computation of EECA rests on the coincidence of actuality and price knowledge only for the more frequently purchased items and not for all items. It appears reasonable to assume that the consumers — especially the Responders, who are presumably very price-sensitive — have good price knowledge for items they buy frequently.

2.3.3 Heterogeneity in Household-Level Parameters

The parameters \(\alpha^h, \alpha^\mu, \{\psi^0, \psi^1\}\), and the probability that \(z(h) = 0\) and are all estimated for each household in a hierarchical Bayes framework. For the Responders, the distribution of the parameters \(\alpha^h\) (in the extreme major-trip model, equations (5), (6)) across households is assumed to be a mixture of multivariate Gaussians\(^5\):

\[
p_{\alpha|z=1}(\alpha; \Theta_{\alpha|z=1}) = \sum_{k=1}^{k_{\alpha}} \pi_{\alpha k} k N(\alpha; \mu_{\alpha k}, \Sigma_{\alpha k}). \tag{13}
\]

We use \(N(x; \mu, \Sigma)\) to denote the value of the density function at \(x\) for the Gaussian distribution with mean \(\mu\) and covariance matrix \(\Sigma\). \(\Theta_{\alpha} \equiv \{\pi_{\alpha k}, \mu_{\alpha k}, \Sigma_{\alpha k}\}_{k=1}^{k_{\alpha}}\) is the set of parameters defining the Gaussian mixture density. For the parameters \(\{\alpha^h\}\), we have a Gaussian mixture with parameters \(\Theta_{\alpha h|z=1}\). For the parameters \(\{\psi^0, \psi^1\}\) which determine the nature of the interpolation between the extreme major and fill-in trip probabilities, we have the Gaussian mixture parameter \(\Theta_{\psi|z=1}\). For the Non-Responders, we have corresponding sets of heterogeneity parameters \(\Theta_{\alpha^h|z=0}, \Theta_{\alpha^\mu|z=0}\) and \(\Theta_{\psi|z=0}\).

It is worthwhile to stress that it is very important to allow for across-household variation in the parameter values, particularly in the category saliences. Our informal conversations with consumers who switch stores because of featured promotions suggest that (a) any one consumer will anchor

\(^4\)It is easy to show that the first order approximation is:

\[
CA^h_{ic} \approx \log(|B_c|) + \sum_{b \in B_c} u^h_{ic}(b, s) \times p^h_{ic}(b | i_c, s).
\]

See section 2.3.3 for an explanation of the notation. Therefore, the category attractiveness is approximately the sum of the items’ utilities, weighted according to the purchase probabilities. Hence, variations in the prices that define the utilities of infrequently purchased items do not affect the sum very much.

\(^5\)We use the Gaussian mixture model of heterogeneity at several points in this paper. The mean and variance of each component will denoted by \(\mu\) and \(\Sigma\), subscripted appropriately. It is important to note that this usage of \(\mu\) is distinct from that in equation (2), where it refers to the store choice probability on an extreme major trip.
his/her store choice on only a small set of categories, and (b) the set of anchor categories will vary from consumer to consumer. For instance, a college student’s choice may be driven by promotions in the frozen foods category, whereas a young mother’s choice may be driven by the baby food and diaper categories. Because each Responder is sensitive to feature advertisements only in a small number of categories, we would expect that (on average) each category would appeal to just a limited subset of consumers (and not appeal to many others). A model which does not contain within-household differences will capture only the category effect average across all households, which will of course be only a weak effect typically. The model of across-household variation we consider in the hierarchical Bayes model here is flexible enough to accommodate substantial across-household differences.

This almost completes our specification of our model for mapping the influence of feature advertising on store choice. We say “almost” because we have yet to state explicitly the expressions for category attractiveness term $CA_{tc}^h$ and the incidence probability $p_t^h(i_c|s)$, which go into the computation of $EECA_{sc}^h(t)$ (all three of these constructs were first introduced in section 2.2). Our expressions for these two terms follow what is commonly posited in the literature for the purchase incidence model.

The purchase incidence model for any one category $c$ is taken to be as given below. Because this purchase incidence decision is viewed here to be independent across categories, we consider each category in isolation. Let $X_{tc}^h(s)$ be the linear predictors of the deterministic utility for purchase incidence. Let the corresponding vector of coefficients be $\beta_{tc}^h$. The probability that a consumer will purchase in a certain category is posited to be given by a binomial probit model:

$$p_t^h(i_c|s) = \Phi((\beta_{tc}^h)^T X_{tc}^h(s))$$

The $\Phi(\cdot)$ is the cumulative distribution function for the standard Gaussian variate. The quantities
that define the values of the elements of $X_{tc}^{h}(s)$ on any shopping occasion are taken to be:

\[ 1 = \text{an intercept term} \]

\[ \{d_{s'}\}_{s' \in S} = \text{a dummy variable taking value 1 if } s' = s \text{ and value 0 otherwise.} \]

\[ I_{tc}^{h} = \text{inventory (introduced in equation (1)), which is inversely related to whether} \]
\[ \text{a trip is an extreme major trip as opposed to an extreme fill-in trip.} \]

\[ INV_{tc}^{h} = \text{our estimate of the level of the household’s inventory in the product category} \]
\[ c, \text{ just before embarking on trip } t. \]

\[ CA_{tc}^{h} = \log(\sum_{b' \in B_{c}} \exp u_{tc}^{h}(b', s)), \text{ a measure of the overall attractiveness of the} \]
\[ \text{store’s offerings in the category. This is the “inclusive term” used in nested} \]
\[ \text{logit models.} \]

Kahn and Schmittlein (1992), Bucklin and Lattin (1991) and Bell and Lattin (1998), among others, demonstrate that the base probability of purchases in a category is lower on a fill-in trip than on a major shopping trip. We use $I_{tc}^{h}$, introduced earlier in our description of the store choice model, as a measure of the extent to which the trip is major or fill-in.

The term $u_{tc}^{h}(b', s)$ is the deterministic component of the utility that household $h$ associates with buying brand-size $b$ in category $c$ at store $s$ on trip $t$. We posit that the probability that the shopper chooses brand-size $b$ is

\[ p_{tc}^{h}(b|s) = \frac{\exp u_{tc}^{h}(b, s)}{\sum_{b' \in B_{c}} \exp u_{tc}^{h}(b', s)}. \]

The set $B_{c}$ is the set of brand-sizes in category $c$. Let $u_{tc}^{h}(b, s)$ be modeled as a simple linear function of the vector of predictor variables (such as price, promotion type etc), so that we can write

\[ u_{tc}^{h}(b, s) = (\gamma_{c}^{h})^{T} Y_{tc}^{h}(b, s). \]

$\gamma_{c}^{h}$ is the vector of utility coefficients for household $h$ in category $c$. The following quantities, measured for brand-size $b$ in store $s$ at the time of trip $t$ constitute the elements of the vector.
\( Y_{tc}^h(b, s): \)

\[
\{d_{b'}\}_{b' \in B_c} = \text{dummy variables taking value 1 if } b' = b \text{ and value 0 otherwise,}
\]

\( LB_{tc}^h(b) = 1 \) if this brand-size was purchased during the last purchase in this category, and 0 otherwise,

\( PRICE_{tc}(b, s) = \) price of brand-size as it appears on the store shelf (this price may be below the store’s “regular price” for the item),

\( PROMO_{tc}(b, s) = \) price discount (regular price less the promotion price) as a fraction of the regular price,

\( FEAT_{tc}(b, s) = 1 \) if the brand-size was promoted through a feature advertisement, and

\( DISP_{tc}(b, s) = 1 \) if the brand-size was promoted through a special display in the store.

Note that \( \gamma_c^h \) is a

\[ \text{household-specific vector of parameters.} \]

The price and promotion inputs listed above are based on actual values rather than on the Bayesian belief distribution like in the store choice model. In the store choice model, we used a Bayesian belief distribution because except for what is advertised, the price and promotion levels in the store are unknown to the consumer. However, at the time of the purchase incidence and brand-size choice decisions, the consumer is already in the store and clearly has access to the actual values of price and promotion levels of all items. Therefore it is reasonable to use actual values of price and promotions variables as inputs to the brand-size choice model.

A very important point to note is the inclusion of \( FEAT \) in the model for brand-size choice. This is intended to allow us to consider the transaction building role of feature advertising separate from its traffic building role (which is represented directly in our store choice model).

In Kamakura and Russell (1989), the distribution of the \( \gamma_c^h \) across households has \( k_{\gamma_c} \) point masses, where \( k_{\gamma_c} \) is the number of latent segments. Rossi and Allenby (1993) have pointed out correctly that such a model does not permit within-segment heterogeneity. They propose instead that the distribution of the \( \gamma_c^h \) be Gaussian. While this takes care of the heterogeneity problem, it views the utility coefficient vectors as Gaussian deviations from a common mean, in effect assuming that there is only one segment. Our model posits that the distribution of the coefficient vector across households is a mixture of \( k_{\gamma_c} \) Gaussian densities:

\[
p_{\gamma_c}(\gamma_c; \Theta) \equiv \sum_{k=1}^{k_{\gamma_c}} \pi_{\gamma_c} N(\gamma_c; \mu_{\gamma_c}, \Sigma_{\gamma_c}).
\]

Here \( \Theta = \{\pi_{\gamma_c}, \mu_{\gamma_c}, \Sigma_{\gamma_c}\} \) are the parameters of the Gaussian mixture. The \( \{\pi_{\gamma_c}\} \) may be interpreted as the relative sizes of the \( k_{\gamma_c} \) consumer segments. This specification for unobserved heterogeneity was used by Allenby, Arora and Ginter (1998).
To model heterogeneity in $\beta^h_c$ across consumers, we similarly assume that the distribution of $\beta^h_c$ across households is a Gaussian mixture density:

$$p_{\beta_c}(\beta_c; \Theta_{\beta_c}) \equiv \sum_{k=1}^{k_{\beta,c}} \pi_{\beta_kc} N(\mu_{\beta_kc}, \Sigma_{\beta_kc}),$$

where

$$\Theta_{\beta_c} = \{\pi_{\beta_kc}, \mu_{\beta_kc}, \Sigma_{\beta_kc}\}_{k=1}^{k_{\beta,c}}.$$

### 2.4 Estimation Procedures and Challenges: Distinguishing between Responders and Non-Responders

Hierarchical Bayes models where the generating prior consists of a mixture of distributions have appeared often enough in the literature that the difficulties and challenges they present have been noted and appreciated and suggestions have been made to address those challenges. What is unusual and different about the central store choice model presented in this paper is not that the generating prior is a mixture density but that it is a mixture of *incomparable* models. This kind of estimation problem has been discussed in the Markov Chain Monte Carlo in only one other paper that we know of (Carlin and Chib 1995). Since the situation in this paper is more complex than their situation, it is worth discussing the basic ideas involved.

In hierarchical Bayes models, a household-level parameter is estimated (approximately) by taking household-level maximum likelihood estimates and shrinking them toward the mean of the population-level distribution. This would be easy if the population-level distribution is known but this too is an unknown to be estimated. The population-level distribution is estimated from the tentative household-level parameters. This estimated population-level distribution is then used to re-estimate the household-level parameters which are then used to re-estimate the population-level distribution. The cycle repeats until stochastic equilibrium is attained. This basic idea can be used if all the households followed comparable models. If the households follow *incomparable* models the procedure needs to be revised. In our store choice model, there are some households which are Responders and some which are Non-Responders and the household-level vectors are incomparable in that they enter into their respective store-choice models in different ways with different model specifications. As a result, one cannot use the household-level parameters directly to revise the population-level density. There is one population density for the Responders and another population density for Non-Responders with domains being the respective (incomparable) household-level parameters. One first needs to know which of the two shopping segments a household belongs
to and that household’s parameter vector should be used to update the population density for only segment and not the other segment. To determine which of the two shopping segments a household belongs to, we do the following: (1) Assume that the household is a Responder and come up with a household-level parameter estimate by shrinking toward the population mean for the Responder segment. (2) Now assume that the household is a Non-Responder and come up with a household-level parameter estimate by shrinking towards the population mean for the Non-Responder segment. (3) Assign the household to either the Responder segment or the Non-Responder segment according to which of these assumed segments (along with their respective household-level parameter estimates) fits the household’s observed store choices better. What makes the estimation process more difficult than MCMC processes in the usual hierarchical bayes settings, is that households can shift between segments and this affects sample and sample-size of households being used to update the population-level distribution.

Care has to be taken as to the exact sequence of updates. Not all update sequences guarantee ergodic behavior for the Markov Chain and convergence to a distribution that is equivalent to the Bayesian posterior density. What we presented above is only the basic intuition for how we collectively estimate the shopper segment for each household, the household-level parameter and two population-level densities. A more careful specification of the exact steps is given in Appendix B.

3 Data, Exploratory Analyses and Model Results

In this section, we describe the source data used to estimate the models presented in Section 2. We present some exploratory analyses and measurements in an attempt to characterize the phenomena we are studying. Finally, we present the main empirical findings from the store choice model.

3.1 Description of Source Data

A consumer panel dataset provided by the market research firm Information Resources Inc feeds the empirical model building, testing, and evaluation in this work. The dataset tracks the grocery purchases of 548 panelists in five stores in a large metropolitan area of the U.S. The data are collected over a two year period. The dataset records purchases made in the following twenty-four categories of UPC-coded products: analgesics, bacon, barbecue sauce, bath tissue, butter, cat food, cereal, coffee, cookies, crackers, detergents, eggs, fabric softener, hot dogs, ice cream, paper towels, peanut butter, pizza, snack foods, soap, soft drinks, sugar, toothpaste and yogurt. The
dataset also contains the store choice and total dollar amount spent on each shopping trip, even if
the panelist did not purchase in any of the twenty-four product categories on that trip. Information
on purchases other than in these twenty-four categories is available only through the total bill for
the trip. In particular, we have no direct information on purchases of produce and meat; these
categories account for a considerable fraction of stores’ revenues and are prominently featured in
the stores’ weekly advertising. This is an important limitation of the dataset because the twenty-four
categories together account for only a fourth of the panelists’ total purchases in the stores. Still,
assuming that there are no strong interactions between the missing product categories and the ones
that we are studying here, the empirical analyses in this paper will contribute valuable insights
on feature advertising in the categories we do include in our analyses. A further limitation of the
dataset is that it records only those trips and purchases made in the five stores. It is likely that the
panelists fill some of their grocery product needs in other stores. Our data providers tell us they
have tried to minimize such purchase leakage by including only those panelists who shop very little
outside of these stores. On the retailer’s side, the dataset reflects all the promotions actions we
are studying. For any given store, item and week, we know the item’s shelf price, and the type of
in-store advertising and feature advertising (if any) for it. For each store, for each item and for each
week, the dataset records whether or not a printed advertisement appeared in that week’s flier for
that item.

To get some feel for the extent of store-switching in the data, let us look at some summary
statistics. By “store-switching,” we mean a consumer visiting a store different from the store
he/she visits most frequently. For the purposes of this analysis, we consider only the “large” trips,
defined to be those trips on which the dollar volume of purchases was greater than the median
volume for that consumer. For example, if a consumer’s frequencies of store-visits to the five stores
in the panel are 0.15, 0.10, 0.70, 0.02, 0.03, then the consumer’s store switching probability is
$1 - 0.70 = 0.30$. The consumer’s probability of switching away from the two most frequented
stores is $1 - (0.70 + 0.15) = 0.15$. Consider now the across-consumer distribution of the store
switching probability. It turns out that median (50%ile) value of this distribution is 0.04. (This
means that half the consumers shop in stores other than their principal store less than 4% of the
time, i.e. a majority of consumers are extremely store-loyal.) The 70%ile for this distribution is
0.19, the 80%ile is 0.28 and the 90%ile is 0.38. Consider now the across-consumer distribution
of probability of switching away from the two most frequented stores. The 90%ile value for this
distribution is 0.04 and the 95%ile value is 0.09. These numbers show consumers overwhelmingly
shop in at most two stores and most of them shop in just one store.
3.2 Estimates for the Purchase Behavior Models

We now present the principal estimation results for the store choice model described in this paper. Estimation is done using Markov Chain Monte Carlo by iteratively drawing from a sequence of conditional distributions. The conditional distributions and mechanisms used to draw from these conditional distributions are described in detail in an appendix available from the authors.

Recall that estimates of the parameters from the purchase incidence model and brand-size choice model enter into predictor variables of the store choice model. Therefore, the results presented here are based on estimates from all three models. Because of limitations imposed by the scope of the dataset in terms of the product categories and number of panelists it embodies, we are unable to estimate models in the full generality described in Section 2 of this paper. However, actual retailers, who are the intended implementors and users of these models will not generally face these limitations. Following are two simplifications forced by the limitations of the dataset:

1. The focus of the empirical work in this paper is more on store choice than in-store behavior. Of the twenty-four categories, five show only a small level of featuring. Therefore, these five categories are excluded from the analysis. One would a priori expect that the less intensely featured categories have a smaller influence on store choice. The nineteen categories that are included in the analysis are:

   Bacon, Bath Tissue, BBQ Sauce, Butter, Cat Food, Cereal, Coffee, Cookies, Crackers, Detergent, Fabric Softener, Hot dog, Ice cream, Paper Towels, Peanut Butter, Pizza, Snack Chips, Soft Drinks, and Yogurt.

2. All the Gaussian heterogeneity models considered are restricted to have zero within-segment correlations. This does not imply that an individual’s value for one parameter gives no information about his/her value for another parameter. Because the population heterogeneity distribution is a mixture distribution composed of Gaussian densities, any two parameters will in general be statistically correlated even if the within-segment Gaussian densities accommodate no dependence. We do not consider within-segment covariances for the Gaussian-distribution components because the number of parameters becomes too large for the choice dataset of so small a scope as the one used in this paper to estimate with sufficient statistical power. For instance, for the bacon category’s brand choice model discussed below, we would need fifty-five additional parameters for each segment. Estimation which considered full covariance matrices produced confidence intervals for the covariance terms that were too wide to be of practical use.
We now present some empirical results to communicate the flavor of the statistical computations in this paper and some of the consequences of the model choices. Conceptually, the estimation for the models in this paper is as follows:

A. Brand Choice Model: A Gaussian-mixture heterogeneity model is estimated for each category.

B. Purchase Incidence Model: The category attractiveness term (inclusive value) based on the brand-choice model is a predictor in the purchase incidence model, which again is estimated under a Gaussian-mixture parameter heterogeneity across households.

C. Store Choice Model: The category attractiveness and the purchase purchase incidence probabilities for the various categories are information elements used as predictors in the store choice model, which again is estimated under a Gaussian-mixture parameter heterogeneity across households.

The focus of this paper is on store-choice behavior in general and on the effect of feature advertising on store choice in particular. Accordingly, our focus in this presentation of empirical results will be on the last of the three sets of models.

**Store Choice Model Results**

We now present the estimated results from the store choice model. Recall that at the individual level we have the following parameters:

1. For all consumers: the zero-order store-choice process logit parameters for the extreme fill-in trip. ($\alpha^h_{\phi s}$)

2. For all consumers: The logistic regression coefficients that map the inventory estimate to be a convex combination of store choice probabilities on an extreme fill-in trip and an extreme major trip on a particular time occasion. ($\psi^h_0, \psi^h_1$).

3. For Non-Responders only: the zero-order store-choice process logit parameters for the extreme major trip. ($\alpha^h_{\mu s}$)

4. For Responders only: the logit store-choice process logit parameters that allow for the effect of feature advertising on the store choice on an extreme major trip. ($\alpha^h_{\mu s}, \alpha^h_{\mu c}$).

The estimation procedure identified a total of eleven segments with the segment-sizes being 0.180, 0.170, 0.120, 0.051, 0.046, 0.110, 0.088, 0.029, 0.099, 0.037, and 0.061. Bayes Factor
scores computed using the Newton-Raftery (1994) method were used for selecting the number of segments.

**Results for Extreme Fill-in Trips**

We do not present the estimates of the $\alpha_{\phi_s}$ terms themselves because they are not directly interpretable. (Refer to equation (4) for the definition of the $\alpha_{\phi_s}$. The subscript $s$ denotes a store, which may be one of A, B, C, D or E.) Rather, we report the characteristics of $\phi_{s}^{h}$, the zero-order store visit probabilities for extreme fill-in trips within each segment, taking into account both the mean of and the within-segment variance of the store logit coefficients. See Equation (4) for the mapping from $\alpha_{\mu_{\phi_s}}$ to $\phi_{s}^{h}$. The within-segment means and standard deviations of the zero-order store visit probabilities is reported in Table 3.

From Table 3, the general characteristics of the eleven segments are more clear: Segments 1, 2, 3, 4 and 5 are very close to “loyal” to stores A, B, C, D, and E respectively. Consumers in segments 1, 2, 3, 4 and 5 are “non-switchers” for the most part while the remaining segments are switchers (Grover and Srinivasan 1992). Consumers in segment 6 are “switchers,” who switch between stores B and E. Consumers in segment 7 are “switchers,” who switch between stores A and C with a preference for store A. Consumers in segment 8 are “switchers,” who switch between stores A and C with a preference for store C. Consumers in segment 9 form a diffuse group which appears to shop in stores A, B, D and E. Notice that all of the logit store-preference coefficients have high within-segment standard deviation. Consumers in segment 10 are Responders who switch between stores A and C. Consumers in segment 11 are Responders who switch between stores B and E.
Table 4: For the Non-Responder segments 1 through 9: The zero-order store choice probabilities $\mu_s^h$ (store $s=A,B,C,D,E$) for the extreme major trips.

(We know them to be Responders because their store choice decision for extreme-major trips is influenced by feature advertising. We will say more about this later in this section.) The pattern of switching in segments 6, 7, 8, 10 and 11 is not surprising: Stores A and C are located in close proximity to each other as are stores B and E. Restrictions on the part of our panel data provider preclude our reproducing a map showing the locations of these five stores in this document.

**Results for Extreme Major Trips (Non-Responders)**

We again do not present the estimates of the $\alpha_{\mu_s}^h$ terms themselves because they are not directly interpretable. (Refer to equation (5) for the definition of the $\alpha_{\mu_s}$. We present instead in Table 4 the store choice probabilities for an extreme major trip for nonresponders, using the same style as in Table 3. Notice that the set of stores that each segment visits predominantly is roughly the same as for the extreme fill-in trips, though the store visit probabilities themselves are quite different. Notice that store-loyalty levels are considerably higher for the extreme-major trips than for extreme fill-in trips. As argued earlier, for the Non-Responder segments, high-volume shopping are more routinized than are low-volume shopping trips, thereby making the store choice decision on an extreme-major trip more stable and less variant over time.

**Results for Major versus Fill-in trips**

Some interesting results come from the parameters that map from $I_t^h$ to $M_t^h$. Recall that $I_t^h$ is an estimate of the household’s total inventory of grocery goods just before trip $t$, measured in dollar value. See section 2.1 for details. Recall also that $M_t^h$ is the degree to which the trip is a major shopping trip versus a fill-in trip. The relation between the two is given by equation (3) and the parameters $\psi_0$ and $\psi_1$. Recall that in the inventory flow equation (1), there is a parameter $\eta$ which weighs more recent transactions heavier than more distant transactions. As we explain in
Table 5: For both Responders and Non-Responders: the parameters $\psi_0$ and $\psi_1$. A * for an estimate indicates a coefficient statistically significant at the 95% level

<table>
<thead>
<tr>
<th>Segment</th>
<th>$\psi_0$ mean</th>
<th>$\psi_1$ mean</th>
<th>$\psi_0$ std. dev.</th>
<th>$\psi_1$ std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.831*</td>
<td>0.023*</td>
<td>0.429</td>
<td>0.022</td>
</tr>
<tr>
<td>2</td>
<td>-1.053*</td>
<td>0.050*</td>
<td>0.415</td>
<td>0.036</td>
</tr>
<tr>
<td>3</td>
<td>-0.862*</td>
<td>0.023*</td>
<td>0.473</td>
<td>0.020</td>
</tr>
<tr>
<td>4</td>
<td>-0.845*</td>
<td>0.034</td>
<td>0.329</td>
<td>0.024</td>
</tr>
<tr>
<td>5</td>
<td>-0.824*</td>
<td>0.036</td>
<td>0.426</td>
<td>0.036</td>
</tr>
<tr>
<td>6</td>
<td>-1.031*</td>
<td>0.072*</td>
<td>0.391</td>
<td>0.047</td>
</tr>
<tr>
<td>7</td>
<td>-1.075*</td>
<td>0.040*</td>
<td>0.366</td>
<td>0.026</td>
</tr>
<tr>
<td>8</td>
<td>-1.144*</td>
<td>0.047</td>
<td>0.325</td>
<td>0.034</td>
</tr>
<tr>
<td>9</td>
<td>-1.207*</td>
<td>0.057</td>
<td>0.380</td>
<td>0.042</td>
</tr>
<tr>
<td>10</td>
<td>-0.983*</td>
<td>0.060*</td>
<td>0.395</td>
<td>0.042</td>
</tr>
<tr>
<td>11</td>
<td>-0.862*</td>
<td>0.041*</td>
<td>0.386</td>
<td>0.028</td>
</tr>
</tbody>
</table>

The text following that equation, the value of $\eta^h$ is chosen to minimize the variance of the estimates, across all shopping trips, of the inventory immediately after a trip. The median $\eta^h$ so computed across all households turns out to be 0.986 with a standard deviation of 0.038. Because the standard deviation is so small, $\eta$ was held fixed across all households to 0.986 in the computation of predicted inventory-level $I^h_t$.

Table 5 gives for each of the eleven segments, the across-household mean and standard deviation of the parameters $\psi_0$ and $\psi_1$. The coefficient values of $\psi_1$ reported here are standardized coefficients in that they are the model estimates of $\psi_1$ multiplied by the standard deviation of the inventory variable across all trips across all households.

The coefficients $\psi_1$ are positive meaning that the greater the current inventory the lower the likelihood that the store choice is for a major trip.

**Effect of Feature Advertising on Responders’ Extreme-Major Trip Store Choices**

Table 6 reports the expected values and the segment-level standard deviations of $\alpha^h$ (see equations (5)–(6)) for the Responder segments in the estimated store choice model. Recall that the two Responder segments are segments 10 and 11.

From the segment-size numbers given in section 3.2, we can see that the two Responders
segments are found to comprise 9.8% of the 548 panelists in the sample. This is well below trade publication estimates (see FMI 2001, for example) which put the figure in the range 15-25%. We suspect that our estimate is lower because of limitations of the dataset. As we pointed out earlier, our dataset lacks data on produce and meat. These categories are believed to be important drivers of store choice; therefore, their omission may diminish the estimate of the fraction of Responders in the sample to the extent that there are Responders only in the omitted categories and not Responders vis a vis the nineteen product categories we consider here.

As can be seen from the coefficients of the store intercepts for the two Responder segments in Table 6, Segment 10 is dominated by those who switch between the store “A” (whose intercept is set to 0) and the store “C”. The Responder Segment 11 consists mainly of consumers who switch among “B,” “E,” and to a smaller extent “D.” The relative sizes of the two Responder segments are 38% and 62%. As we mentioned earlier, this switching pattern is consistent with spatial locations of the five stores. Table 6 also reports the segment-level means of the $\alpha_{hc}$ coefficient estimates for the two Responder segments. Unfortunately, the coefficients $\alpha_{hc}$ are not interpretable by themselves. This is because the scales of the category attractiveness values are very different across categories. Therefore, the relative values of the coefficients are driven to a significant degree by the differences in scales of the category attractiveness measures. Therefore, the value of $\alpha_{hc}$ for each of the nineteen product categories as shown in Table 6 is multiplied by the standard deviation of the category attractiveness (EECA) value. This is comparable to the kind of rescaling produced in “standardized regression” coefficients. An asterisk marks coefficients that are statistically significant at the 5% level.

**Demographic Differences Between Responders and Non-Responders**

The MCMC procedure used in this paper gives estimates of $z$ for each household so that we can classify each household as a Responder versus a Non-Responder. The dataset comes with some demographic data on each household which can be used to study differences between households classified as Responders and households classified as Non-Responders. The demographic variables considered are family size, household income, ethnicity and binary variable indicating whether or not the household subscribes to a daily newspaper. We performed a simple logistic regression to predict the Responder versus Non-Responder status of a household as a function of the demographic variables. The response variable is 1 if the household is classified as a Responder and 1 if classified as a Non-Responder. The only demographic variable that we found to be a statistically significant predictor is the household size. Larger households are less likely to belong to the Responder segment presumably because they are time constrained.
<table>
<thead>
<tr>
<th>Category/Store</th>
<th>Seg’t 10</th>
<th>Seg’t 11</th>
<th>Seg’t 10</th>
<th>Seg’t 11</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>mean</td>
<td>std. dev.</td>
<td>std.dev</td>
</tr>
<tr>
<td>Store A</td>
<td>0.000*</td>
<td>0.000*</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Store B</td>
<td>-2.371*</td>
<td>3.702*</td>
<td>0.994</td>
<td>1.335</td>
</tr>
<tr>
<td>Store C</td>
<td>-0.431*</td>
<td>0.821*</td>
<td>0.188</td>
<td>0.255</td>
</tr>
<tr>
<td>Store D</td>
<td>-2.270*</td>
<td>2.722*</td>
<td>0.818</td>
<td>1.034</td>
</tr>
<tr>
<td>Store E</td>
<td>-4.342*</td>
<td>3.301*</td>
<td>2.145</td>
<td>1.345</td>
</tr>
<tr>
<td>Bacon</td>
<td>0.091</td>
<td>0.067</td>
<td>0.052</td>
<td>0.021</td>
</tr>
<tr>
<td>Bath Tissue</td>
<td>0.072</td>
<td>0.093</td>
<td>0.043</td>
<td>0.028</td>
</tr>
<tr>
<td>BBQ Sauce</td>
<td>0.018</td>
<td>0.028</td>
<td>0.008</td>
<td>0.014</td>
</tr>
<tr>
<td>Butter</td>
<td>0.022</td>
<td>0.016</td>
<td>0.010</td>
<td>0.006</td>
</tr>
<tr>
<td>Cat Food</td>
<td>0.044</td>
<td>0.036</td>
<td>0.023</td>
<td>0.012</td>
</tr>
<tr>
<td>Cereal</td>
<td>0.107*</td>
<td>0.117*</td>
<td>0.044</td>
<td>0.054</td>
</tr>
<tr>
<td>Coffee</td>
<td>0.109*</td>
<td>0.116*</td>
<td>0.049</td>
<td>0.047</td>
</tr>
<tr>
<td>Cookies</td>
<td>0.104*</td>
<td>0.104*</td>
<td>0.050</td>
<td>0.046</td>
</tr>
<tr>
<td>Crackers</td>
<td>0.080</td>
<td>0.114*</td>
<td>0.039</td>
<td>0.035</td>
</tr>
<tr>
<td>Detergent</td>
<td>0.107</td>
<td>0.138*</td>
<td>0.034</td>
<td>0.051</td>
</tr>
<tr>
<td>Fabric Softener</td>
<td>0.063</td>
<td>0.052</td>
<td>0.024</td>
<td>0.024</td>
</tr>
<tr>
<td>Hot dog</td>
<td>0.100*</td>
<td>0.048</td>
<td>0.034</td>
<td>0.016</td>
</tr>
<tr>
<td>Ice cream</td>
<td>0.060</td>
<td>0.113*</td>
<td>0.028</td>
<td>0.060</td>
</tr>
<tr>
<td>Paper Towels</td>
<td>0.038</td>
<td>0.016</td>
<td>0.015</td>
<td>0.006</td>
</tr>
<tr>
<td>Peanut Butter</td>
<td>0.045</td>
<td>0.014</td>
<td>0.016</td>
<td>0.006</td>
</tr>
<tr>
<td>Pizza</td>
<td>0.141*</td>
<td>0.025</td>
<td>0.064</td>
<td>0.012</td>
</tr>
<tr>
<td>Snack Chips</td>
<td>0.107*</td>
<td>0.129*</td>
<td>0.042</td>
<td>0.057</td>
</tr>
<tr>
<td>Soft Drinks</td>
<td>0.074</td>
<td>0.054</td>
<td>0.026</td>
<td>0.029</td>
</tr>
<tr>
<td>Yogurt</td>
<td>0.071</td>
<td>0.109*</td>
<td>0.044</td>
<td>0.037</td>
</tr>
</tbody>
</table>

Table 6: The store choice model with two Responder segments. The table shows the expected value of $\alpha^h_{\mu}$ over all the households in each segment. The raw coefficient for each of the nineteen product categories is multiplied by the standard deviation of the category attractiveness (EECA) value; this is comparable to the kind of rescaling produced in “standardized regression” coefficients. An asterisk marks coefficients that are statistically significant at the 5% level.
3.3 Model Assessment by Cross-Validation

To study the predictive validity of feature advertising, we compare the accuracy of the prediction rule obtained for the Responders in the eleven-segment solution with the prediction rule obtained from an eleven-segment solution with only Non-Responders. The test sample consists of 10% of the store-choice observations and is never used in the estimation procedure. For this validation test, we consider only those households which are classified as Responders by the store choice estimation procedure. There are 48 such households. The number of trips in the test sample for these households is 471. To predict a household’s store choice, we use the max-score rule: we estimate the store choice probabilities for any given occasion using equation (5) and propose the store with the highest choice probability as the predicted choice. For the model which allows no Responders, the prediction accuracy is 57.1%. For the prediction rule based on the Responders model, which uses information from feature advertising, the prediction accuracy is 67.8%. The difference is significant at the 1% level. (We use a paired-difference test.) The improvement in accuracy is statistically significant and substantial. Perhaps including more product categories, particularly those in meats and produce, would improve the predictive accuracy even more.

4 Applications in Decision Support Systems

The core store choice model developed in this paper allows one to relate managerial decisions on feature advertising to household-level store choice probabilities and, by aggregation, to store traffic levels which in turn affect store-wide sales levels. Therefore, the model can be used to decide which items to advertise and what price promotions level to announce in store fliers. In this section, we consider two examples illustrating how the core store model can be used in a decision support context.

4.1 Promotion Elasticity of Store Choice

An important piece of information that the core store choice model can offer for the retailer is an estimate of the featuring elasticity of store choice probability. As an illustration, consider the question, What is the percentage increase in the probability of choosing store “C” when all the prices in a certain category are decreased by 1% and advertised? Assume that competing stores’ marketing levels are held unchanged. The elasticity would be a useful measure of the category’s salience in the store choice decision. No simple formula exists for computing this elasticity. We
compute this number by simulation, and by focusing on the Responder segments. The results are reported in the second column in Table 7.

Table 7 also shows the ranks of categories according to actual level of featuring by store C in the dataset. If retailer C is feature-advertising to maximize a draw of consumers to Store C, then we would expect that the categories with the highest store choice elasticities would be the ones that would be feature-advertised the most. A comparison of the actual ordering of feature advertising levels with the store choice elasticities is interesting. The soft drinks category is a highly promoted and featured category. It is striking that the soft drinks category has rather low store choice elasticity. We would like to suggest an explanation: The heavily featured categories are very likely those categories where most stores in the market will promote very frequently. The result is that discounts in that category become commonplace and the category’s importance in the store choice decision (statistically speaking) diminishes because there is very little variance across stores. An examination of the competing stores’ pricing and advertising for the soft drinks category suggests that this explanation is indeed plausible: every store in our dataset promotes a major soft-drink in most weeks.

A variation on our measure of store choice elasticity

In our analysis above, we considered the situation where every item was discounted by 1% and the discount announced via feature advertising in the store’s weekly flier. Strictly speaking, a retailer would probably not consider such a promotion; usually only select items from a category are promoted. We could recompute the elasticity assuming that only the single brand with the highest market share is promoted. While this would reflect what commonly occurs in categories like Bacon or Soft Drinks where most of the market is occupied by a very small number of brands, it does not reflect what happens in categories like Cereal where even the most dominant brand occupies a very small share of the category. In such categories, it is common to find several brands promoted in any one week. To approach what happens in real world promotions practice, we consider the case where the top brands collectively accounting for at least 20% of category share are promoted. As before, we consider a price cut of 1% announced via feature advertising in the weekly store flier. The values under this different operationalization are reported in Table 8. We preserve the ordering of the previous table to facilitate comparison. One can see the rank ordering with these revised elasticities is quite close to the rank ordering from the previous elasticities. Indeed, the two sets of elasticities are highly correlated: the rank correlation is 98%. 
<table>
<thead>
<tr>
<th>Category</th>
<th>Elasticity of Choice</th>
<th>Rank by Actual Feature Advertising Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pizza</td>
<td>0.203</td>
<td>4</td>
</tr>
<tr>
<td>Hot dog</td>
<td>0.181</td>
<td>6</td>
</tr>
<tr>
<td>Cereal</td>
<td>0.173</td>
<td>11</td>
</tr>
<tr>
<td>Cookies</td>
<td>0.148</td>
<td>14</td>
</tr>
<tr>
<td>Coffee</td>
<td>0.131</td>
<td>15</td>
</tr>
<tr>
<td>Snack Chips</td>
<td>0.121</td>
<td>9</td>
</tr>
<tr>
<td>Ice cream</td>
<td>0.120</td>
<td>7</td>
</tr>
<tr>
<td>Bath Tissue</td>
<td>0.090</td>
<td>3</td>
</tr>
<tr>
<td>Detergent</td>
<td>0.090</td>
<td>10</td>
</tr>
<tr>
<td>Soft Drinks</td>
<td>0.082</td>
<td>1</td>
</tr>
<tr>
<td>Bacon</td>
<td>0.061</td>
<td>2</td>
</tr>
<tr>
<td>Paper Towels</td>
<td>0.060</td>
<td>5</td>
</tr>
<tr>
<td>Butter</td>
<td>0.045</td>
<td>8</td>
</tr>
<tr>
<td>Yogurt</td>
<td>0.031</td>
<td>12</td>
</tr>
<tr>
<td>Crackers</td>
<td>0.030</td>
<td>13</td>
</tr>
<tr>
<td>BBQ Sauce</td>
<td>0.030</td>
<td>17</td>
</tr>
<tr>
<td>Fabric Softnr</td>
<td>0.024</td>
<td>18</td>
</tr>
<tr>
<td>Catfood</td>
<td>0.023</td>
<td>19</td>
</tr>
<tr>
<td>Peanut Butter</td>
<td>0.023</td>
<td>16</td>
</tr>
</tbody>
</table>

Table 7: The second column reports estimates of the elasticity of store choice for the store “C.”. The categories are sorted in order of decreasing elasticity. The third column ranks each category in terms of actual feature advertised discounts: In the Soft Drink category, advertised price discounts are the heaviest relative to other categories.
<table>
<thead>
<tr>
<th>Category</th>
<th>Elasticity of Choice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pizza</td>
<td>0.046</td>
</tr>
<tr>
<td>Hot dog</td>
<td>0.060</td>
</tr>
<tr>
<td>Cereal</td>
<td>0.043</td>
</tr>
<tr>
<td>Cookies</td>
<td>0.054</td>
</tr>
<tr>
<td>Coffee</td>
<td>0.044</td>
</tr>
<tr>
<td>Snack Chips</td>
<td>0.038</td>
</tr>
<tr>
<td>Ice cream</td>
<td>0.042</td>
</tr>
<tr>
<td>Bath Tissue</td>
<td>0.028</td>
</tr>
<tr>
<td>Detergent</td>
<td>0.021</td>
</tr>
<tr>
<td>Soft Drinks</td>
<td>0.030</td>
</tr>
<tr>
<td>Bacon</td>
<td>0.022</td>
</tr>
<tr>
<td>Paper Towels</td>
<td>0.014</td>
</tr>
<tr>
<td>Butter</td>
<td>0.012</td>
</tr>
<tr>
<td>Yogurt</td>
<td>0.014</td>
</tr>
<tr>
<td>Crackers</td>
<td>0.012</td>
</tr>
<tr>
<td>BBQ Sauce</td>
<td>0.011</td>
</tr>
<tr>
<td>Fabric Softnr</td>
<td>0.005</td>
</tr>
<tr>
<td>Catfood</td>
<td>0.009</td>
</tr>
<tr>
<td>Peanut Butter</td>
<td>0.010</td>
</tr>
</tbody>
</table>

Table 8: The column of numbers reports estimates of the elasticity of store choice for the store “C.” under the revised promotion experiment where only the top 20% of the brands are promoted and advertised.
4.2 Optimal Portfolios of Categories to Promote via Feature Advertising

The results from the previous section suggest that if Store C were forced to pick just one category where the top 20\% of the brands were promoted, then the best categories to pick would be Hot Dog. In this section, we generalize this kind of optimizing and consider now the case where the store is to pick the best $k$ categories to promote via feature advertising. If the total number of categories is $C$ then there are \( \binom{C}{k} = \frac{C!}{(C-k)!k!} \) possible subsets of size $k$ to consider. This number can be very large for large values of $C$ even if the value of $k$ is not very large. Therefore, it may be combinatorially infeasible to examine all possible subsets. Hence, one may need to consider heuristic search methods. The problem of best category subset selection we are considering here is similar to the problem of best variable subset selection that comes up in regression. This latter problem is well studied and there are several solution methods as described in Miller (1992). In regression with correlated variables, the predictive value of a variable depends on what other variables are already in the variable subset under examination. We see a similar characteristic in the context of feature advertising optimization: the increase in store visits created by a category being advertised depends on what other categories are being advertised.

One heuristic search method that is found to perform well in similar combinatorial optimization problems is a “greedy” local search method that works as follows. The goal is to maximize the average probability of shopping in store C. Start with a random subset of size $k$.

1. Deletion: Drop feature advertising in the category where dropping advertising least diminishes the average store choice probability for Store C. Note that there are $k$ candidate categories to consider for dropping. For each candidate that is dropped, we need to recompute the objective function, which is an expensive computation that involves simulations based on the store choice model and the heterogeneity distributions.

2. Addition: Add feature advertising in the category where adding advertising most increases the average store choice probability for Store C. There are $C - k$ candidate categories to consider for adding, not counting the category that was dropped in the previous step. For each candidate that is added, we need to recompute the objective function.

We iterate through the two steps above until convergence, which happens when the Addition step above merely reverses the Deletion step. In our experience with this algorithm for optimizing feature advertising bundles, the algorithm takes at most approximately $9k$ iterations to converge. Since each iteration takes a total of $C$ function evaluations (each of which can be very time consuming),
the total number of function evaluations is less than $9Ck$, often much less than this figure. This is a far smaller number than the $\binom{C}{k}$ function evaluations it would take for an exhaustive search.

It is important to note that the above algorithm produces only a local optimum that may vary according to the random subset which was used at the commencement of the iterations. However, in our experiments, different starting values subsets frequently produced the same final subset. On the occasions that the final subsets were different, the corresponding objectives function values were not substantially different. A further encouraging fact is that when we tried the $k = 2$ case where exhaustive search was feasible, the local search produced the same subset as the one with exhaustive search.

As an illustration, we give results for the $k = 5$ case. We have 19 product categories, so $C = 19$. In this instance, we take feature advertising in a category to mean that the top brands accounting for at least 20% share are given a 1% advertised price-cut by Store C, similar to the second scenario we considered in the previous subsection. Exhaustive search would require evaluating $\binom{19}{5} = 11628$ distinct subsets. The greedy local search algorithm took only 16 iterations, for a total of just 285 subset evaluations. The best subset of categories to feature-advertise was found to be: Cereal, Snack Chips, Pizza, Cookies and Hot dog. The combined elasticity of simultaneous promotions in these five categories turned out to be 0.321. To get some idea as to the impact of such optimizing decision support systems, we computed the store choice elasticity to promoting in 40 randomly chosen category subsets of size $k = 5$ and found the average to be 0.151. This means that the optimizing subset achieves a lift that is just over double the lift achieved with a random portfolio, which is a substantial improvement.

If the number of categories $C$ becomes very large, then even approximation heuristics like ours become difficult to execute, not just because $9Ck$ becomes large but because each function evaluation grows more expensive. In such situations, our best suggestion would be to take the best set $k$ highest categories for feature-advertising to be simply the $k$ categories with the largest advertising elasticities of store choice. In our example, with $k = 5$, applying this heuristic would mean that we would pick the top 5 categories from Table 8: Hot dog, Cookies, Pizza, Coffee, Cereal. The combined elasticity of simultaneous promotions in these five categories was found to be 0.308. Comparing with the figure 0.151 obtained for a randomly chosen subsets, we see that the model can be very useful in picking subsets even if heuristic subset optimization is not feasible.
5 Conclusions and Summary

This paper develops a model of consumer shopping behavior that considers the effect on feature advertising. Because consumers vary in their cost-saving strategies, only some consumers use feature advertising to decide which store to visit. Therefore, any study of the impact of feature advertising should focus on this Responder consumer segment. We discussed the determinants of the choice behaviors of Responders and the role of retailer promotion programs. This paper studies the effect of feature advertising on store choice through a model of choice at the level of the consumer.

This approach forces the researcher to explicitly consider important phenomena in purchase behavior (for example, that price expectations and perceptions of basket needs have specific roles). A methodological contribution of our paper is the ability to provide price expectations over items prior to store visit based only on the partial information contained in the store flier.

The traffic-building power of feature advertising arises because of the Responder consumer segment. The research and managerial issues we raise above can be addressed only if (a) we can identify the Responders (in the panel) and (b) examine their characteristics at the household level (rather than at the cross-sectional level). Both of these requirements present methodological difficulties. The chief methodological advance of this work is in its ability to segment the consumer panelists into Responders and Non-Responders. This is a challenging problem in statistics because we do not observe the consumer’s use of feature advertisements, and hence have to infer the consumer’s decision strategy from store choice decisions. We accomplish this through a careful predictive model-selection scheme implemented in a hierarchical Bayes framework and estimated by the method of Markov Chain Monte Carlo. One of our substantive finding is that the Responder consumer segment consists of only about 10% of the shopper population. This being such a small fraction, the methodological problem of identifying who the Responders are and measuring the extent to which different categories drive store choice becomes critically important.

Using our estimated model of store choice, we report feature-advertising elasticities of store choice for various categories. An interesting finding is that the categories that are feature-advertised most heavily are not the categories where advertising in actuality has the highest impact on store choice. The calibrated model provides useful insights at the level of the consumer. The model also has the potential to help managers choose the product categories to promote as illustrated in our section on using the model in decision support systems.
Appendix A

In this appendix, we derive the expression given in equation (11) for the conditional expectation of a function over the Gaussian kernel density estimate. From equation (10), the conditional density is

\[
f(P_{sc} | P_{sc}^{(k)} = P_{sc}^{(k)}) = \frac{1}{N} \sum_{k=1}^{N} \frac{1}{\sqrt{2\pi d|H_N|}} \exp \left( -\frac{1}{2} ([P_{sc}^{(k)} - P_{sc}^{(u)}] H_N^{-1} ([P_{sc}^{(k)} - P_{sc}^{(u)}]^T) \right)
\]

The expectation of any function \( g(P_{sc}) \) over this conditional density can be written as

\[
E(g(P_{sc})|P_{sc}^{(k)} = P_{sc}^{(k)}) = \frac{1}{N} \sum_{k=1}^{N} \int \frac{1}{\sqrt{2\pi d|H_N|}} \exp \left( -\frac{1}{2} ([P_{sc}^{(k)} - P_{sc}^{(u)}] H_N^{-1} ([P_{sc}^{(k)} - P_{sc}^{(u)}]^T) \right) g(P_{sc}) dP_{sc}^{(u)}
\]

Classical nonparametric theory dictates that for the estimate (9) to be consistent, the bandwidth matrix \( H_N \) has to diminish with \( N \) (the theoretical best rate is \( O(N^{-1/5d}) \), see Scott (1992) for details). Under the same conditions, for large \( N \), each integral in the numerator’s sum above can be approximated accurately as:

\[
\int \frac{1}{\sqrt{2\pi d|H_N|}} \exp \left( -\frac{1}{2} ([P_{sc}^{(k)} - P_{sc}^{(u)}] H_N^{-1} ([P_{sc}^{(k)} - P_{sc}^{(u)}]^T) \right) g(P_{sc}) dP_{sc}^{(u)} \approx \frac{1}{\sqrt{2\pi d|H_N|}} \exp \left( -\frac{1}{2} (P_{sc}^{(k)} - P_{sc}^{(n)}) H_N^{-1} (P_{sc}^{(k)} - P_{sc}^{(n)})^T) \right)
\]

(15)

To understand why this approximation is valid, notice that the convolving function in the integral is proportional to a Gaussian. As \( H_N \) gets smaller, the Gaussian gets concentrated about the mode \( P_{sc}^{(u)} n \). Therefore the convolution integral can be approximated as simply the integrand evaluated at the mode. Substituting equation (15) in (14) and simplifying, we can estimate the conditional expectation as

\[
E(g(P_{sc})|P_{sc}^{(k)} = P_{sc}^{(k)}) = \frac{\sum_{n=1}^{N} w_n g([P_{sc}^{(k)} - P_{sc}^{(n)}])}{\sum_{n=1}^{N} w_n},
\]

\[
w_n = \exp \left( -\frac{1}{2} (P_{sc}^{(k)} - P_{sc}^{(n)}) H_N^{-1} (P_{sc}^{(k)} - P_{sc}^{(n)})^T) \right).
\]

39
Appendix B

Consider first the special situation where all the households are known to belong to Responders. In this case, estimating all the parameters is simple. Given values of \( \{\alpha^h, \psi^h\}_{h=1}^H \), we can generate variates from the conditional marginals for \( \Theta_\alpha|z=1 \) and \( \Theta_\psi|z=1 \) (as in previous models, following Bayesian estimation for the Gaussian case). The values of \( \{\alpha^h, \psi^h, \phi^h, \tau^h\}_{h=1}^H, \Theta_\alpha|z=1, \Theta_\psi|z=1, \Theta_\phi|z=1, \) and \( \Theta_\tau|z=1 \) can all be generated by BMHIC sampling. Now consider the case where all the households belong to Non-Responders. We can estimate the parameters exactly the way we did in the previous case, if we recognize that the Non-Responders are Responders with \( \{\alpha^h_c = 0\}_{c=1} \).

How do we estimate the store choice model parameters in the general case, where the consumer pool consists of both Non-Responders and Responders? A simple observation crucial to our method for estimation of the store choice model is this. If we knew the values of \( \{z(h)\}_{h=1}^H \) then we could estimate separate the Responders from the Non-Responders, and reduce the estimation problem to the two special cases described above. The only missing link then is the posterior distribution of \( z(h) \) conditional to the data and all the other parameters. If we had this, then we could use Gibbs sampling to generate samples from the joint posterior of all parameters. We next show how to generate samples from the conditional marginal densities for \( \{z(h)\}_{h=1}^H \).

First, some simplifying notation: Let \( \xi^h_{z=1} \) denote the set of store choice model parameters for household \( h \), given that \( h \) is a Responder. Let \( \xi^h_{z=0} \) denote the set of parameters for household \( h \), given that \( h \) is a Non-Responder. Let \( \pi_{z=1} \) be the generating prior probability that a panelist is a Responder. Finally let \( p(\xi^h_{z=0} ; \Theta_\xi|z=0) \) and \( p(\xi^h_{z=1} ; \Theta_\xi|z=1) \) be the generating priors for \( \xi^h_{z=0} \) and \( \xi^h_{z=1} \) respectively. The values of \( \xi^h_{z=0} \) and \( \xi^h_{z=1} \) are non-comparable in the sense that they are drawn from different parameter spaces and therefore have different interpretations. Following Carlin and Chib (1995), model selection can be done by considering Gibbs transitions on the joint parameter space, \( \{\xi^h_{z=0}, \xi^h_{z=1}, z(h)\} \). Straightforward algebra shows that the conditional marginal posteriors
in this joint space are:

\[
p(\xi_{z=0}^h|\cdot) \propto \begin{cases} 
L(D^h|\xi_{z=0}^h, z(h) = 0) p(\xi_{z=0}^h; \Theta_{\xi|z=0}) & \text{if } z(h) = 0 \\
p(\xi_{z=0}^h; \Theta_{\xi|z=0}) & \text{if } z(h) = 1
\end{cases}
\]

\[
p(\xi_{z=1}^h|\cdot) \propto \begin{cases} 
L(D^h|\xi_{z=1}^h, z(h) = 0) p(\xi_{z=1}^h; \Theta_{\xi|z=1}) & \text{if } z(h) = 0 \\
L(D^h|\xi_{z=1}^h, z(h) = 1) p(\xi_{z=1}^h; \Theta_{\xi|z=1}) & \text{if } z(h) = 1
\end{cases}
\]

\[
p(z(h) = 1|\cdot) = \frac{p(\pi_{z=1}; n_0 + \sum_{h=1}^H 1_{z(h)=1}, n_0 + \sum_{h=1}^H 1_{z(h)=0})}{L(D^h|\xi_{z=1}^h, z(h) = 1) + \pi_{z=1} L(D^h|\xi_{z=1}^h, z(h) = 0)}
\]

Therefore, the odds ratio of Non-Responder versus Responder is the ratio of store choice likelihoods in the two sections, adjusted according to the segment sizes. One might worry that that \(p(D^h|\cdot, \gamma_h, z(h) = 1)\) overstates the “true” likelihood because Responders have more parameters and more flexibility of fit. While this is true in the frequentist MLE setting, it is not a problem in the Bayesian case where we are sampling \(\alpha_h\) from its entire posterior (and not just its maximum). Further details are given in a technical appendix available from the authors.
References


Efron, Bradley (1982), *The Jackknife, the Bootstrap, and other Resampling Plans*, Philadelphia: SIAM.


Geweke, J. (1992), “Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of


Rossi, Peter E. and Greg M. Allenby (1993), “A Bayesian approach to estimating household pa-


