Advertising, Learning, and Consumer Choice in Experience Good Markets: An Empirical Examination

Daniel A. Ackerberg*

UCLA and NBER
March 27, 2002

Abstract

This paper empirically analyzes different effects of advertising in a non-durable, experience good market. A dynamic learning model of consumer behavior is presented in which we allow both "informative" effects of advertising and "prestige" or "image" effects of advertising. This learning model is estimated using consumer level panel data tracking grocery purchases and advertising exposures over time. Empirical results suggest that in this market, advertising’s primary effect was that of informing consumers. The estimates are used to quantify the value of this information to consumers and evaluate the welfare implications of an alternative advertising regulatory regime.

---

*Economics Dept, UCLA, Los Angeles, CA 90095 (ackerber@econ.ucla.edu). This paper is a revised version of the second and third chapters of my 1997 doctoral dissertation at Yale University. Many thanks to my advisors, Steve Berry and Ariel Pakes, as well as Lanier Benkard, Russell Cooper, Sara Fisher Ellison, Gautam Gowrisankaran, Sam Kortum, Mike Riordan, John Rust, Roni Shachar, and many seminar participants for advice and comments. I thank the Yale School of Management for gratefully providing the data used in this study. Financial support from the Cowles Foundation in the form of the Arvid Anderson dissertation fellowship is acknowledged and appreciated. All remaining errors are my own.
1. Introduction

Theoretical work in economics has long been concerned with different influences of advertising on consumer behavior. Marshall (1919) praised "constructive" advertising, described as advertising that conveys economically relevant information to consumers. On the other hand, he termed the "incessant iteration of the name of a product" as "combative" advertising, and criticized the "social waste" of such behavior. More recently, economists have developed formal models of advertising. Stigler (1961), Butters (1977), and Grossman and Shapiro (1984) examine models where firms send advertising messages to explicitly inform consumers of their brand's existence or observable characteristics. In contrast to this explicit information, Nelson (1974), Kihlstrom and Riordan (1984), and Milgrom and Roberts (1986) analyze models in which firms producing non-durable experience goods use advertising to implicitly signal information on their brand's experience characteristics (e.g. unobserved quality). In these equilibria, brands with higher unobserved quality advertise more and consumers rightfully interpret these high advertising levels as a signal of this higher quality.

Stigler and Becker (1977) and Becker and Murphy (1993) examine models in which a brand's advertising level interacts with consumption in the consumer's utility function. In this model, consumers may simply derive more utility from consuming a more advertised good (analogous to the excess utility some might derive from dining in a "prestigious" restaurant). One can make similar arguments where consumers derive direct utility from advertising content, e.g. images or personalities. In contrast to the above "informative" effects of advertising, we term these "prestige" or "image" effects of advertising. As these prestige and image effects involve advertising in itself changing demand for a brand, the framework provides a way of capturing the ideas behind Marshall's "combative" advertising and Galbraith's (1958) "persuasive" advertising that is consistent with rational consumers and utility maximization. Evidence of such effects might be Coca-Cola and Pepsi television advertising. We doubt that this level of advertising would be optimal if its sole purpose was to provide product information to the very few consumers who do not already know the existence or characteristics of the brands.

One implication of this theoretical literature is that the way in which advertising affects consumers
impacts the functioning of a market. Advertising that provides information on search or experience characteristics is likely to have different implications on market structure, evolution, and performance than advertising which creates prestige or image associations that give direct utility to consumers. Unfortunately, theory cannot tell us which of these effects exist or predominate in a particular market. In certain markets, casual empiricism may suggest an answer, e.g. Coke and Pepsi. However, there is a wide range of markets where the answer is not clear. Past empirical literature addressing this question has suffered from a variety of limitations. Telser (1964) and Boyer (1974) correlate advertising levels and measures of profitability at the industry level. Though interesting, their identifying hypothesis, that informative effects should reduce entry barriers and profitability while non-informative effects should raise them, suffers from acknowledged endogeneity problems. Benham (1972) and Milyo and Waldfogel (1999) rely on unique natural experiments. Nelson (1974) includes empirical work that suggests signaling content of advertising, but his methods cannot formally measure or separate different effects. Resnik and Stern (1978) examine actual advertisements to assess informational content. However, information that a product exists or implicit signaling information need not be embodied in explicit verbal or visual content.

1.1. Empirically Distinguishing Different Effects of Advertising

This study follows Ackerberg (2001) in capitalizing on consumer level panel data to distinguish and measure different effects of advertising. Our data follows consumers' purchases and television advertising exposures for a newly introduced brand of Yogurt over a 15 month period. The goal is to determine whether these advertisements provided product information to consumers, generated Becker-like prestige or image effects, or both. Ackerberg (2001) addresses this question using a reduced form empirical approach, looking for a differential effect of these advertisements on experienced and inexperienced consumers of the brand (experienced consumers being those who have tried the brand at some point in the past). Since experienced consumers presumably already know of the brand’s existence and its effects. One example is entry. If advertising purely provides information, ability to advertise may decrease “informational” barriers to entry in an industry (see e.g. Tirole (1988) pg. 289). On the other hand, prestige effects might increase barriers to entry by creating product differentiation and market power.
observable and unobservable characteristics, he argues that they should not be affected by exposures to informative advertising\(^2\). In contrast, he hypothesizes that Becker-like prestige or image effects of advertising should generally affect both inexperienced and experienced users of the brand\(^3\). Simple reduced form discrete choice models indicate that the advertisements did affect consumers who had never experienced the brand of yogurt before but did not affect experienced consumers. He concludes that these advertisements provided information.

The present study applies a similar identification argument from a more structural perspective. To more rigorously examine these informational arguments, we formally model consumer information, introducing a model of consumer behavior that explicitly includes both informative and prestige effects of advertising. Our model is similar to Eckstein, Horsky, and Raban's (1989) dynamic learning model of experience goods with the addition of these two effects of advertising. In each time period, dynamically optimizing consumers choose whether to purchase a non-durable, experience good. Consumers start the model with imperfect information on the brand's characteristics. They learn about these characteristic both through consumption of the brand and through informative advertising. Our Becker-like prestige or image effect of advertising enters directly in the utility function, influencing utility independently of beliefs on inherent product characteristics\(^4\). A related model is estimated by Erdem and Keane (1996). They also extend Eckstein, et. al. to include informative advertising, but examine the demand implications of this single effect and do not distinguish different effects of advertising\(^5\).

---

\(^2\)One of the noted exceptions is advertising providing information on changing search characteristics, e.g. price. Price information, however, is not typically mentioned in the television advertisements for non-durables like those considered here. See Ackerberg (2001) for other exceptions and more discussion.

\(^3\)The idea here is that if a consumer obtains an extra \(z\) utils from consuming a product that is associated (by advertising) with a particular image, seeing that ad will increase the utility obtained from consuming the product by \(z\) regardless of whether he has purchased in the past. Clearly there is a bit of speculation in formulating these intangible image and prestige effects, so we try to be as general as possible in specifying them. On the other hand, a key to empirically distinguishing these effects from informative effects is the assumption that they do not interact in the utility function with past consumption. An example of such an interaction is a consumer who gets less prestige utility from current consumption of a brand the more he has consumed the brand in the past.

\(^4\)We stress that these prestige/image effects constitute completely rational behavior on the part of consumers. Termining these “persuasive” effects of advertising might be somewhat of a misnomer, as our consumers are not somehow persuaded or fooled by advertising into making bad purchase decisions.

\(^5\)There are other significant differences between the two models. One is the extent of consumer heterogeneity. Our model studies one brand and allows consumers to differ in both initial and final (post-information) valuation of the brand. In Erdem and Keane, there is no heterogeneity in what is learned, so consumers all converge to the same beliefs. On the other hand, they are able to examine learning and advertising for multiple (8) brands. The models also differ in
The simplest representation of the important empirical components of our model is as follows. Suppose a consumer purchases a brand if the utility he expects to obtain from consuming the brand is greater than some threshold \( k \), i.e. \( i \)

\[
E[U(\pm a) | a] > k
\]

The utility function \( U \) contains \( \pm \), representing the brand's inherent characteristics (e.g. calories, fat content, taste), and \( a \), some measure of what the consumer knows about the brand's advertising. The expectation is over \( \pm \) as the consumer is uncertain about the brand's characteristics. \( a \) enters in two places into this expected utility. First, it directly enters the utility function. This is our prestige or image effect of advertising - advertising influencing utility given inherent product characteristics. Secondly, the expectation over \( \pm \) is conditioned on \( a \). This is our informative effect of advertising - we allow advertising to "tell" the consumer something about the brand's characteristics \( \pm \). As consumption of the brand also provides information to the consumer on the brand's characteristics, the model implies that informative advertising impacts the expected utility of inexperienced consumers more than that of experienced consumers. On the other hand, our prestige or image effect of advertising affects utility regardless of whether a consumer is experienced or not. This distinction is what separately identifies these two effects of advertising in the structural model.

Formalizing this model involves specifying the process through which informative advertising affects a consumer's information set. There are a number of different types of information advertising can provide: explicit information on product existence or observable characteristics, or signaling information on experience characteristics. It would be optimal to write down and estimate a consumer model including all these possible informative effects. Unfortunately, such a model would likely be computationally intractable, and more importantly, these separate informative effects would be hard, if not impossible, to empirically distinguish given our dataset. We therefore choose just one informative effect to include in our structural model, that of signaling. Reasons for this choice include: (1) the

the way that informative advertising is modelled and in the policy analysis that is performed. They examine alternative firm advertising strategies while we measure the value of information in advertising.
recent focus on signaling arguments in the theoretical literature to explain the lack of explicit information in many television advertisements, (2) some casual empirical evidence from Ackerberg (2001), and (3) convenience and flexibility in computation and estimation. Given the necessity of making such a choice, it is very important to note that this empirical work does not take a stand on which types of informative effects of advertising are actually occurring in our market. However, we believe that these different informative effects of advertising should in some sense be observationally equivalent in our data: all tend to affect inexperienced rather than experienced consumers. As a result, we feel that our conclusions regarding significance or insignificance of our informative and prestige effects would not substantially change if we had instead modeled one of the other informative effects. In summary, we interpret a statistically significant signaling effect of advertising not as empirical support for signaling per se, but as support for the more general hypothesis that advertising is providing some kind of product information to consumers.

1.2. Motivation for the Structural Model

There are important advantages of this structural approach relative to the reduced form models of Ackerberg (2001). If consumers learn from consumption of a brand (and the data suggest they do), we expect to see discrete (and likely persistent) changes in consumer behavior after consumption experiences. More specifically, if consumers obtain idiosyncratic information from consumption, we might expect prior experience and the resulting accumulation of information to generate relatively higher variance (across consumers) in experienced consumers' behaviors (e.g. some consumers find out they like the brand, some find out they do not). This increased dispersion in behavior is not captured in standard discrete choice models where explanatory variables, e.g. \( \text{prior experience} \), shift means and not variances. This contrasts with our structural learning model, which does accommodate such dispersion by allowing heterogeneous consumer tastes for the brand that are not realized (learned) by a consumer until after having experienced the brand\(^{6}\). Not only will ignorance of this dispersion

\(^{6}\)In the reduced form models, adding a random coefficient on a dummy variable “prior experience” might be able to partially replicate this dispersion. However, such models begin to look a lot like the myopic structural models used in this paper.
be inefficient, but it can potentially generate spurious results\textsuperscript{7}. This illuminates the need to consider structural models in empirical studies of information.

A second major advantage of the structural approach is that it allows for interesting policy analysis that is simply not possible with reduced form results. If, for example, advertising provides consumers with information, we would like to know the value of this information. In order to compute such a value, we need to be able to adjust optimal consumer behavior when the source of information is eliminated. With a structural model this is possible, unlike reduced form models that would suffer from the Lucas (1971) critique. We stress that, unlike our main empirical conclusions, the welfare analysis we perform is probably highly dependent on our choice to model informative advertising as a signaling effect. Though this limits the applicability of the welfare results, we feel that it is still an interesting and enlightening exercise.

1.3. Summary of Results

Estimates of our structural learning model support two main conclusions. First, we can easily reject the hypothesis of perfect information. The data suggest that consumers do learn from their consumption experiences with the brand. Second, we find a strong, positive informative effect of advertising and an economically and statistically significant prestige effect of advertising. This supports the reduced form conclusion that the advertisements in this data primarily affected consumers through the provision of information. Under the strong assumption that this is in fact signaling information, our policy analysis indicates that the value of this information to consumers is significantly less than the resources spent on advertising. This at least suggests that advertising signaling may be a very inefficient way of transferring information. Section 2 introduces our general model of consumer behavior and Section 3 describes the data used in this study. Section 4 details our empirical specification and presents our results. In Section 5 we perform our welfare experiments and Section 6 concludes.

\textsuperscript{7}For example, consider a situation where both experienced and inexperienced consumers have $E[U(\delta, a) \mid a]$'s (EU's) distributed around zero (assume consumers purchase if $EU > 0$), but experienced consumers EU’s have more dispersion (a higher variance) (see Figure 1). In this case, a burst of prestige advertising that shifts all consumers EU’s up by a certain amount will induce a higher proportion of inexperienced consumers than experienced consumers to purchase. Without conditioning on this increased dispersion, one would incorrectly conclude that this advertising relatively affects inexperienced consumers.
2. The Model

Consider a consumer who in each time period $t$, observes prices, $p_{it}$, and advertising intensities, $a_{it}$, of a newly introduced non-durable experience good. Advertising intensity refers to some measure of the number of advertisements for the brand that consumer $i$ is exposed to in period $t$, perhaps divided by units of possible exposure time (e.g. TV watching time). Note that prices are allowed to vary across both consumers and time. It is assumed that the good is non-durable enough so that a brand purchased at $t$ is completely consumed before $t + 1$.

After observing prices and advertising intensities in a given period, the consumer decides whether to purchase one unit of the brand or the outside alternative. Consumers are assumed to make this discrete choice to maximize their expected discounted sum of future utilities conditional on their information set at $t$:

$$\max_{c_t(I_{it}) \geq 2} \sum_{\tau=t}^{\infty} \bar{\delta}^{\tau-t} U_{ic_{\tau}j \mid it}$$

where $c_t \in \{1, 2\}$ is the consumer's choice at $t$ (2 represents the outside alternative) and $\bar{\delta}$ is the per-period discount factor.$^8$

As is now relatively common in the empirical analysis of differentiated products, we take a Lancasterian, characteristics-based approach to consumer theory, assuming the utility a consumer derives from a brand is a function of the brand's characteristics and the consumer's tastes for these characteristics. Specifically, we assume the utility consumer $i$ obtains from purchase and consumption of brand $j$ in period $t$ is:

$$U_{it} = \gamma_i + \mu_1 p_{it} + \xi_{it+1} + \mu_2 m_{it}^o + \epsilon_{it}$$

$\gamma_i$ represents consumer $i$'s tastes for the observable characteristics of the new product. This will generally be a function of the product's observable characteristics and the consumer's known tastes for these characteristics. $\epsilon_{it}$ represents idiosyncratic, time-varying shocks to the utility the consumer

---

$^8$Note that we consider an infinite horizon problem. Though consumers' lives are in fact finite, the time frame of the empirical work is approximately weekly, so the number of periods will be very large. Also note that since a consumer's information changes through time, the maximization problem is over a sequence of choice functions mapping future information sets into choices.
derives from consuming that are known prior to the purchase decision. Though we defer its formal
definition until later, m_{it} is a measure of what consumer i currently knows about how much the
brand is advertising. Its entry into the utility function will represent our image or prestige effect of
advertising. For now, suppose the utility from consuming the outside alternative, U_{i2t}, in any period
is 0.

The term ±_{t+1}, which we call \"experience utility\", captures the experience nature of the good. It is
a scalar measure of the utility that consumer i derives from brand characteristics that are not directly
observable to him (i.e. experience characteristics). It is dated t + 1 because in contrast to the other
elements of the utility function, it is not necessarily known to the consumer at the time of purchase.
For food products, ±_{t+1} might capture how the brand actually tastes to the consumer (conditional
on its observed characteristics that enter \( \bar{\varepsilon}_i \)). Although ±_{t+1} is not observable before purchase, it is
observed if the good is purchased and consumed at t because total utility is realized and all other
components of the utility function are known. Thus, in the simplest case where ±_{t+1} is constant over
time we have a \"one-consumption\" learning process. In this case, after the consumer purchases and
consumes the brand once, he observes ±_{t+1} and knows its value for all future t:

As in Eckstein et.al. (1988), we allow for a more general learning process in which it may take
more than one consumption to ascertain the experience utility to expect from future consumption of
a brand. Specifically, it is assumed that:

\[ ±_{t+1} = ± + \varepsilon_{it+1} \quad \text{where} \quad \varepsilon_{it+1} \sim \text{iid } N(0, \sigma^2_{\varepsilon}) \] (2.3)

Although ±_{t+1} is realized (observed) by the consumer after consumption, its components, ± and
\( \varepsilon_{it+1} \); are never individually observed. ± is the mean experience utility consumer i obtains from the
brand. \( \varepsilon_{it+1} \) are i.i.d. confounding variables that cannot be distinguished from this mean. In the
case of food products, variance in \( \varepsilon_{it+1} \) may result from variation in product quality, combination
with other products, the existence of different flavors of a brand that the consumer must learn to
optimize over, or even different moods or situations at time of consumption\(^9\). In contrast to the i.i.d.
\(\theta_{t+1}, \pm\) is persistent over time. It is thus beneficial for the consumer to use information contained
in observed \(\pm_{t+1}\)'s to learn about its value. In the degenerate case where \(\theta^2 = 0\), we have the
one-consumption learning process described above where \(\pm\) (and thus \(\pm_{t+1}\) at) is learned after one
consumption experience. In the non-degenerate case, consumption and subsequent realization of \(\pm_{t+1}\)
does not exactly reveal \(\pm\), but it does provide information about it. This information acquisition will
be consistently modeled in a Bayesian learning framework.

In a similar formulation, we assume that consumers' observed advertising intensities, \(a_{it}\), follow
the process:

\[
a_{it} = a + \epsilon_{it} \quad \text{where} \quad \epsilon_{it} \sim \text{iidN}(0; \theta^2) \tag{2.4}
\]

where \(a\) is the mean advertising intensity of the brand. Deviations in \(a_{it}\) around \(a\) may be caused
by variation in consumers' television or reading habits or variation in where or when a brand is
advertised\(^{10}\). Although consumers do not directly observe a brand's mean advertising intensity \(a\),
we allow them to be interested in it for two reasons: (1) Possible prestige, image or status effects of
advertising where the consumer, all else equal, obtains more utility from consuming a more advertised
brand or a brand more associated (through advertising) with a particular image, and/or (2) a belief
that firms use a to implicitly signal information on the mean experience utility they obtain from the
brand \(\pm\), as would be the case in a Nelson type signaling equilibrium. In either of these cases, an
optimizing consumer will use observed \(a_{it}\)'s to learn about \(a\). Note that in this model there is no
explicit information about the product obtained through advertisements: consumers are assumed to
know the existence of and the observable characteristics of the brand\(^{11}\).

We consistently model information provided by the observed \(a_{it}\)'s and \(\pm_{t+1}\)'s on the relevant

\(^9\)In all these cases the important thing is that the \(\epsilon_{it}\) are indistinguishable from \(\delta_i\), e.g. the consumer is in a happy
mood, enjoys the product more than usual, but cannot distinguish exactly what component of the extra enjoyment was
due to his mood and what component was due to the product’s experience characteristics.

\(^{10}\)In our empirical work, we generalize to having advertising exposures distributed around an individual specific mean
(i.e. \(a_{it} = a_i + \xi_{it}\)).

\(^{11}\)As discussed in the introduction, complexity and identification issues necessitated the inclusion of only one informative
effect of advertising in our model. There are clearly alternative specifications. Erdem and Keane’s (1996) similar dynamic
model has advertising explicitly informing the consumer on \(\delta_i\). Another alternative would be to allow advertising to inform
consumers of a product’s existence, essentially changing the consumer’s choice set (e.g. like coupons in Leslie (1999)).
unknowns $a$ and $\pm$ as a bivariate Bayesian learning process. In matrix notation, equations (2.3) and (2.4) become

$$
\begin{bmatrix}
\hat{a}_{it+1} \\
\pm_{it+1}
\end{bmatrix} \sim \text{i.i.d. } \mathcal{N}
\begin{bmatrix}
\hat{a} \\
\pm
\end{bmatrix};
\Sigma
$$

where $\Sigma = \begin{bmatrix} \frac{\xi}{2} & 0 \\ 0 & \frac{\xi}{2} \end{bmatrix}$ (2.5)

The assumed diagonality of $\Sigma$ implies that there is no correlation between $\pm_{t+1}$ and $a_t$ conditional on their means. In other words, deviations around mean experience utility due to quality variation, consumption situations, flavors, etc., are assumed uncorrelated with the deviations around mean advertising level due to variation in television watching or brand advertising levels. Appealing to the theory of conjugate distributions (DeGroot (1970)), this equation, along with an initial ($t = 0$) prior on $a$ and $\pm$:

$$
\text{Initial Prior: } \begin{bmatrix}
\hat{a} \\
\pm
\end{bmatrix} \sim \text{i.i.d. } \mathcal{N}
\begin{bmatrix}
m_0 \\
m_0'
\end{bmatrix};
\Sigma_0
$$

(2.6)

generates a learning process in which a consumer's posterior on brand $j$ after a history of observed advertising intensities, $f a_{i1}; \ldots; a_{it}$, and consumption experiences, $f \pm_{i1}; \ldots; \pm_{it}$, is given by\textsuperscript{12}:

$$
\begin{bmatrix}
\hat{a} \\
\pm
\end{bmatrix} \sim \text{i.i.d. } \mathcal{N}
\begin{bmatrix}
m_{it} \\
m_{it}'
\end{bmatrix};
\Sigma_{it}
$$

(2.7)

where:

$$
m_{it} = \frac{i m_{ii}'}{m_{ii}} = (\Sigma_0^{-1} + \hat{\Sigma}_{it}^{-1})^{-1}(\Sigma_0^{-1} m_0 + \hat{\Sigma}_{it}^{-1} z_{it});
$$

$$
\Sigma_{it} = (\Sigma_0^{-1} + \hat{\Sigma}_{it}^{-1})^{-1};
$$

and $K_{it}$ equals the number times the consumer has bought the brand up to period $t$. As the consumer observes an advertising intensity for each brand in each period, the number of observed advertising intensities is $t$. Because of the linearity of the utility function in $a_i$ and $\pm_{i+1}$, setting the initial prior mean on $\pm$ equal to 0 is a normalization. Essentially, we are treating the expected value of the unobserved characteristic as an observable characteristic (i.e., it is part of $a_i$).

\textsuperscript{12}The derivation of the following conjugate result is a fairly simple extension of the derivation for a multivariate normal with equal draws in DeGroot.
\( m_t \) is a weighted average of initial priors and observed realizations of \( \pm_{t+1} \) and \( a_t \). An important result of Bayesian learning is that these posterior means and variances summarize all the consumer's information on \( \pm \) and \( a \). Thus, the current posterior \((m_t, \sigma_t)\) is sufficient to define perceived distributions over both future \( \pm_{t+1} \)'s and \( a_t \)'s as well as future posteriors.

Of particular interest at this point is the composition of the variance matrix of the consumer's initial priors. If the covariance term of \( \sigma_0 \) is zero, then the learning processes on \( \pm \) and \( a \) are independent. On the other hand, a non-zero covariance term indicates a perception by consumers that \( \pm \) and \( a \) are correlated. This links the two learning processes - in this case, observed levels of advertising will not only provide direct information on \( a_j \), but also provide indirect information on \( \pm_t \).

Correlation in initial priors would arise from a belief that advertising is used by firms to signal information on a brand's experience utility. We allow there to be such a signaling equilibrium in which firms set brand advertising levels according to:

\[
a = -\frac{1}{1+\rho_1} \pm
\]

where \( \pm \) is the brand's mean experience utility level over the population. Then, assuming (1) a normal population distribution of \( \pm \) around \( \pm(\pm \sim N(\pm, \frac{1}{2})) \) and (2) a normal prior on \( \pm(\pm \sim N(0, \frac{1}{2})) \) a Bayesian consumer's initial prior variance matrix is:

\[
\text{Initial Prior: } \begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 2 & 31 \\ \rho & \frac{1}{2} & 0 & \rho & 1 & 31 & \frac{1}{2} \\ -4 & 6 & \frac{3}{2} & -\frac{3}{2} & -\frac{3}{2} & 7 \rho \\ \end{bmatrix}
\]

Thus, in a signaling equilibrium where \( -\frac{1}{2} > 0 \), consumers interpret high levels of advertising as a signal of a higher \( \pm \).

In addition to this informative effect of advertising, we accommodate prestige or image effects of advertising by allowing the consumer's current posterior mean on \( a \); \( m^*_t \), to enter directly into the

\[13\] We take the view that consumers are rational in their beliefs, i.e. that a positive covariance term in priors is actually generated by a signaling equilibrium. An alternative interpretation is that we are simply estimating consumer beliefs, without assuming anything about where they come from or whether or not they are correct.
utility function (2.2). As prestige and image effects of advertising are somewhat non-tangible, this term warrants some discussion. One alternative is to interpret this term structurally, i.e. all else equal (in particular, expectations over $x_{t+1}$), consumers simply receive more utility from consuming products that have higher advertising intensities. This is analogous to the structural effect of Becker and Murphy (1993) - in their model, \textbackslash amount of advertising", just like calories or taste, is a product characteristic that confers utility to consumers. Their intuition for the existence of such an effect is similar to that of product characteristics like prestige - in this case, the characteristic might be termed \textbackslash advertising-based prestige\textsuperscript{14}.

A second alternative is to interpret this term as a reduced form representation of more general prestige or image effects of advertising. As an example, suppose consumers obtain higher utility from consuming products that have Kobe Bryant in their advertisements. Structurally this effect might be very discrete - consumers either know or don't know that Kobe Bryant is associated with the product. This knowledge might come from seeing advertisements, talking to other people, or through other means. One would think, however, that the mean level of advertising observed by the consumer would be correlated with this knowledge. As a result, our coefficient on $m_{it}$ should pick up this effect. As another example, image effects may arise from a consumer wanting (i.e. deriving utility from) other people associating them with particular images. Since the amount of other people who are aware of the brand's image association will depend on the amount the brand advertises the association, one can justify $m_{it}$ entering directly into the utility function\textsuperscript{15}. Clearly we are in somewhat murky waters regarding the specification of these image and prestige effects. To partially compensate, in empirical work we try to be as general as possible with the specification. However, we also have to admit the possibility that this representation might pick up misspecification.

Given the learning process as specified in (2.7), we can move back to the consumer's dynamic choice problem. Because the posterior $(m_{it}, \pi_{it})$ is sufficient to determine all the consumer's current

\textsuperscript{14}Note that when we discuss advertising's effect on prestige, we are referring to its effect on a single aspect of prestige (defined as that aspect which is affected by advertising). Other aspects of prestige could be determined by other variables, e.g. packaging, pricing etc. Our hope is that these other aspects of prestige are constant across time and known to consumers, and thus captured by our constant term and consumer random effects.

\textsuperscript{15}We are assuming that each television advertisement portrays the same images, i.e. that advertising copy is the same or similar across commercials. Given that we don't observe advertising copy, this is likely all we can do.
information on $\pm$ and $a$; the sequential maximization problem of (2.1) can be transformed into the following Bellman's equation:

$$
V_i(p_{it}; m_{it}; \$_{it}; ²_{it}) = \max_{c_{it} \in \{1, 2\}} E[U(p_{it}; x_{it}; ²_{it}; m_{it}; \pm_{it+1})
+ \bar{V}_i(p_{it+1}; m_{it+1}; \$_{it+1}; ²_{it+1}) \cdot g(p_{it}; m_{it}; \$_{it}; ²_{it}; c_{it}] \tag{2.8}
$$

where the state space $(p_{it}; m_{it}; \$_{it}; ²_{it})$ contains prices, the current posterior, and time-varying preference shocks $²_{it}$. The expectation is over current period experience utility $\pm_{it+1}$ as well as next period's state. For consumer $i$ with posterior $(m_{it}; \$_{it})$ facing prices $p_{it}$ and shocks $²_{it}$, $V_i(\phi)$ is the perceived expected discounted value of future utilities. This value function has an associated policy function, $c_{it} = c_i(p_{it}; m_{it}; \$_{it}; ²_{it})$, which maps the consumer's current state into the optimal purchase choice.

Ackerberg (1997) provides more details of this Bellman's equation including the corresponding state evolution equations.

To summarize, we have a dynamic model of behavior in which a consumer learns from both consumption and advertising exposures. Purchase and consumption of a brand provides the consumer with direct information on the utility he derives from the brand's experience characteristics. Observed advertising intensities have two effects: First, providing indirect, signaling information on experience characteristics. Second, they are a direct indication of a brand's advertising intensity, which may through image or prestige effects provide a direct utility the consumer.

Unfortunately, the above dynamic model is not analytically solvable. However, we have used numeric solution methods to solve and generate predictions of the model. We detail the more important ones here - Ackerberg (1997) contains a more thorough comparative static analysis. One implication of the learning process is that consumers may change their purchasing patterns over time as a result of new information. The parameters of the learning process determine how long these purchasing patterns will be changing. In a model with no advertising and a one-consumption learning process $(\frac{\theta}{\beta} = 0; [\$_{0}]_{11} > 0)$, purchase patterns change after the first purchase but not thereafter. If there is variance in $\pm_{t+1}$ $(0 < \frac{\theta}{\beta} < 1)$, purchasing patterns do change after future purchases, the extent and
length depending on $\frac{3}{2}$ and $[s_{0}1]$. On the other hand, if there is no learning ($\frac{3}{2} = 1$ or $[s_{0}1] = 0$), we obtain constant purchasing patterns through time.

A second characteristic of the model is that if there is learning, there is a value of information to consumers. Consumers may be willing to experiment with new brands that do not maximize expected current utility in order to obtain information on that brand and make more educated decisions in the future. The extent of this willingness depends on the consumer's discount rate, prior variances, and per-period variances in advertising intensities and experience utility.

The most important implication of the model for the current empirical study concerns the different effects of advertising. Both consumption and informative advertising can provide information to the consumer on $\mathbb{E}$. However, while consumption provides direct information on $\mathbb{E}$, advertising only provides indirect information through consumers' prior beliefs that the two variables are correlated. The more direct information the consumer has obtained through consumption experiences, the less he needs to rely on the indirect advertising information. As a result, all else equal, the more consumption experiences a consumer has had, the less informative advertising will affect his expected utility from consumption. Under one-consumption learning, for example, informative advertising will not affect a consumer after one consumption experience with the brand. In contrast, our direct, prestige or image effect of advertising affects the expected utility of inexperienced and experienced users of a brand equally. This is the behavioral implication that we take to the data to distinguish between the two effects of advertising.

3. The Data

We use consumer-level panel data on grocery purchases to estimate this model. This data, collected by A.C. Nielsen, is commonly referred to as "scanner panel data" because it was recorded by supermarket UPC scanners. In each of two geographically isolated markets (Sioux Falls, South Dakota and

---

16This type of data has primarily been analyzed in the marketing literature (e.g. Guadagni and Little (1983), Podrick and Zafryden (1991), Deighton, et. al. (1994), Russell and Kamakura (1994), McCulloch and Rossi (1994)). With the exception of Erdem and Keane (1996), these studies have used static, more "reduced form", discrete choice models of consumer behavior. As in Erdem and Keane, the studies that examine advertising focus on assessing "how much" advertising affects behavior, not distinguishing "how" it affects behavior.
Springfield, Missouri), shopping trips and purchases of approximately 2000 households at 80% of area supermarkets and drugstores were followed for three years (1986-1988). There is also data on weekly prices at each store, so we essentially know prices on each household’s shopping trips. In addition to containing this extremely detailed data on household purchases over time, A.C. Nielsen TV meters were used to collect information on household TV advertising exposures for about half the households in the last year of the data. We thus know, along with when and what each household bought, when members of the household were potentially exposed to TV advertisements for each brand.

The publicly available Nielsen data contains data on four product categories: ketchup, laundry detergent, soup, and yogurt. Ackerberg (2001) chose to focus on the yogurt data for reasons that are just as relevant for this study. First, the inability to even parsimoniously include inventory behavior and purchase quantity choice in the model suggests the choice of the least durable of the above products. Second, empirical identification in both models relies on distinguishing experienced from inexperienced users of a brand. This generates a serious initial condition problem unless one has data from a product’s initial introduction on the market. The yogurt data includes such a product: Yoplait 150, a lowfat yogurt introduced in April, 1987, about 15 months before the end of the Nielsen data. As computational issues are even more binding here than in that paper, we again focus specifically on Yoplait 150, modeling competing brands in an informationally static and sparse framework.

Table 3.1 gives some summary statistics for the data following Yoplait 150’s introduction. Comparing advertising shares to market shares suggests that it was, at least initially, a heavily advertised yogurt. The large difference in market shares between markets 1 and 2 may be due to the existence of two, high-share, local brands in market 1 and the significant number of manufacturer coupons that seem to have been available in market 2. We urge the reader to consult Ackerberg (1997, 2001) for

---

17When a consumer purchases a product, we observe the exact transaction price. To measure prices when a product wasn’t purchased, other households’ purchases at the same week in the same store were used (prices change weekly). However, for approximately 30% of our observations, extrapolation had to be done from prices in adjacent weeks at the same store. While this is a large proportion and could result in significant measurement error, our focus is on the effects of advertising, not price elasticities.

18Only households whose television viewing was recorded are included both here and in estimation. We also limit the data to shopping trips in supermarkets (rather than drugstores) and those in which $10 or more was spent. The motivation here was to eliminate quick trips to the supermarket for particular items - where Yoplait 150 was likely not part of the choice set.
a more thorough data description, including samples of particular consumer's purchase patterns, an
examination of the time paths of prices, advertising, and sales, and a discussion of data problems
relating to manufacturers coupons\textsuperscript{19} and advertising\textsuperscript{20}.

4. Estimation

We now move to estimation of our model using this data, starting with a more detailed discussion
of our empirical specification, estimation issues and numeric techniques used in solving the dynamic
programming problem. Because the empirical model is fairly complicated and non-linear, we intuitively
discuss how our data should identify the parameters of the learning process. Two general sets of
estimates are presented. The first assumes myopic consumers who learn and update according to
the above model but maximize only current expected utility. Although this model identifies the
parameters of the learning process, it does not require solving the dynamic programming model of
section 2, significantly reducing computational burden. The second set of estimates are of the full
dynamic problem in which consumers are forward-looking in their behavior.

4.1. Empirical Specification

For empirical work, we assume that the time frame of our model is the consumer (i.e. household)-
shopping trip. Specifically, we model consumer i's decision whether or not to purchase Yoplait 150 on
each of their shopping trips t through the 15 months of data. The choice to purchase a different brand
of yogurt is included in the "outside alternative" along with the decision not to buy any yogurt\textsuperscript{21}. We

\textsuperscript{19}Briefly, since we only observe manufacturers coupons that are redeemed, there is an obvious endogeneity problem
using them as explanatory variables. Because of their relative prevalence in market 2, we use a market dummy as a proxy for the "availability of manufacturers coupons". In contrast, we do know when in-store coupons were available (in the data this was only for one week in two stores), so we do include this as an explanatory variable.

\textsuperscript{20}Unfortunately, advertising is only measured in the last year of the data. This leaves about three months when Yoplait
150 was available but advertising was not measured. We use zero advertising exposures for this period. A justification for
this is that for almost three weeks after TV measurement started, there were no Yoplait 150 advertisements observed. This
might suggest that Yoplait did not begin advertising the product until this time. Evidence in Ackerberg (2001) suggests
that alternative treatments of this time period does not affect the identification of different effects of advertising. Another
problem with our advertising variable is some unreliability of TV meters (we eliminated consumers with extremely large
viewing gaps - an indication that their meter may not have been working).

\textsuperscript{21}In preliminary specifications we compared these 2 choice models to 3 choice models (with the choices: Yoplait 150, a
different brand of yogurt, or no yogurt) and obtained similar results. Note that we also completely ignore the number of
yogurts purchased on a particular shopping occasion, avoiding what is in actuality a more complicated discrete/continuous
specify our consumers' single period utility functions as:

$$U_{it} = \begin{cases} 
\lambda_i + \mu_1 p_{it} + \mu_2 sc_{it} + \mu_3 m_{it}^a + \mu_4 m_{it}^b + 2_{i1t} & \text{if Yoplait 150 purchased & consumed at } t \\
\mu_4 p_{oth}^{it} + 2_{i2t} & \text{otherwise (outside alternative)}
\end{cases}$$

The variables $p_{it}; sc_{it};$ and $p^{oth}_{it}$ measure price of Yoplait 150, value of a possible store coupon available for Yoplait 150, and a scalar measure of other yogurts' prices respectively on shopping trip $t$ of consumer $i$. Note that these variables vary over both time and consumers, as supermarkets change prices over time and consumers shop at different supermarkets. The parameters $\mu_1, \mu_2,$ and $\mu_4$ measure marginal effects of these variables on utility.

$\lambda_i$ is modeled as a linear combination of observable consumer characteristics ($y_i$) plus a normally distributed random variable with variance $\sigma^2$. $y_i$ includes a market dummy, the consumer's income and family size, and the number of yogurt, lowfat yogurt, and regular Yoplait purchases made by the consumer in the data prior to Yoplait 150's introduction on the market. The "random effect" component of $\lambda_i$ allows for persistent differences in consumers' known tastes for Yoplait 150 that are not observed by us as econometricians.

To ease computation in both the dynamic programming problem and estimation, $2_{i1t}$ and $2_{i2t}$ are assumed i.i.d. Type 1 Extreme Value deviates. As in a standard discrete choice model, we cannot identify relative levels or variances of the utility function. The lack of a constant term in the outside alternative utility is our additive normalization; the fixed variance of the $2_{i1t}$'s the multiplicative.

In our model, consumers want to learn how much advertising a brand is engaging in. Thus, we define $a_{it}$, consumer i's observed advertising intensity in a given period $t$, as the number of advertisements seen by $i$ between the current ($t$) and previous ($t-1$) shopping trip divided by the amount

---

22 For shopping trip $t$, this is measured as $\min_j \{ (p_{ijt} - \bar{p}_j)/\bar{p}_j \}$, the minimum (over all other brands of Yogurt $j$) percentage current deviation from the average price of that brand.

23 This "presample" purchase data is assumed exogenous to our model, and as might be expected are very good predictors of $\lambda_i$. In Ackerberg(2001) other household characteristics such as ages and sexes were not significant. Note that coefficients on individual observed characteristics of Yoplait 150 (e.g. calories) are not separately identified. $\lambda_i$ represents the sum of the utilities from these characteristics for consumer $i$. 

---

17
of television watched during that period. This controls for the fact that different consumers watch different amounts of television. For the Bayesian updating formulas, we also need to know the per-period variance in \( a_{it} \). This is computed for a particular consumer and time period as a function of the amount of television watched since the last shopping trip and the measured sample variance of advertising intensity in the data. This allows the precision of an advertising observation \( a_{it} \) to increase in the number of hours of television watched between \( t \) and \( t \). Though this directly controls for consumers di®erent in how much television they watch, the data also indirectly suggest that consumers persistently di®er what they watch (we nd statistically signi®cant di®erences in consumers' mean (over time) \( a_{it} \)'s). To accommodate this, we add an additional level of variance to the advertising exposure process. Speci®cally, we assume that the \( a_{it} \)'s are distributed normally around a consumer-speci®c advertising intensity \( a_{i} \) (measuring what consumer \( i \) watches) which in turn are distributed normally around the brand's advertising intensity \( a_{25} \). The variance of \( a_{i} \) around \( a_{25} \), \( \sigma^2_{a} \), is taken directly from the data. In addition, \( a_{i} \) is estimated directly from the data. This implies that we need not estimate the parameter \( \alpha_{0} \), since conditional on the parameters \( \alpha_{1} \) and \( \alpha_{2} \), our signalling equilibrium equation implies \( \alpha_{0} = \alpha_{1} \pm \alpha_{2} \).

4.2. Identi®cation

It is important to discuss how these learning parameters are identi®ed by the data. Identification comes primarily from examining how consumers' purchase behaviors change through time, in particular after the potential acquisition of information from consumption or advertising. If there is no learning, we would see constant (but likely heterogeneous over consumers) purchasing patterns over time (conditional on covariates such as price). With learning, consumption experiences will change a consumer's purchasing patterns. Eventually, everything about the brand is learned and a consumer's purchase parameters change.

\[ a_{it} \]

\( \sigma^2_{a} \)

\( \alpha_{0} \)

\( \alpha_{1} \)

\( \alpha_{2} \)

\( m_{it}^{a} \)

\( a_{i} \)

\( a_{it} \)

\( a_{25} \)

\( \sigma^2_{a} \)

\( \alpha_{0} \)

\( \alpha_{1} \)

\( \alpha_{2} \)

\( m_{it}^{a} \)

\( a_{i} \)

\( a_{it} \)

\( \sigma^2_{a} \)

\( \alpha_{0} \)

\( \alpha_{1} \)

\( \alpha_{2} \)

\( m_{it}^{a} \)

\( a_{i} \)

\( a_{it} \)

\( \sigma^2_{a} \)

\( \alpha_{0} \)

\( \alpha_{1} \)

\( \alpha_{2} \)

\( m_{it}^{a} \)

\( a_{i} \)

\( a_{it} \)

\( \sigma^2_{a} \)

\( \alpha_{0} \)

\( \alpha_{1} \)

\( \alpha_{2} \)

\( m_{it}^{a} \)

\( a_{i} \)

\( a_{it} \)

\( \sigma^2_{a} \)

\( \alpha_{0} \)

\( \alpha_{1} \)

\( \alpha_{2} \)

\( m_{it}^{a} \)

\( a_{i} \)

\( a_{it} \)

\( \sigma^2_{a} \)

\( \alpha_{0} \)

\( \alpha_{1} \)

\( \alpha_{2} \)

\( m_{it}^{a} \)

\( a_{i} \)

\( a_{it} \)

\( \sigma^2_{a} \)

\( \alpha_{0} \)

\( \alpha_{1} \)

\( \alpha_{2} \)

\( m_{it}^{a} \)

\( a_{i} \)

\( a_{it} \)

\( \sigma^2_{a} \)

\( \alpha_{0} \)

\( \alpha_{1} \)

\( \alpha_{2} \)

\( m_{it}^{a} \)

\( a_{i} \)

\( a_{it} \)

\( \sigma^2_{a} \)

\( \alpha_{0} \)

\( \alpha_{1} \)

\( \alpha_{2} \)

\( m_{it}^{a} \)

\( a_{i} \)

\( a_{it} \)

\( \sigma^2_{a} \)

\( \alpha_{0} \)

\( \alpha_{1} \)

\( \alpha_{2} \)

\( m_{it}^{a} \)

\( a_{i} \)

\( a_{it} \)

\( \sigma^2_{a} \)

\( \alpha_{0} \)

\( \alpha_{1} \)

\( \alpha_{2} \)

\( m_{it}^{a} \)

\( a_{i} \)

\( a_{it} \)

\( \sigma^2_{a} \)

\( \alpha_{0} \)

\( \alpha_{1} \)

\( \alpha_{2} \)

\( m_{it}^{a} \)

\( a_{i} \)

\( a_{it} \)

\( \sigma^2_{a} \)

\( \alpha_{0} \)

\( \alpha_{1} \)

\( \alpha_{2} \)

\( m_{it}^{a} \)

\( a_{i} \)

\( a_{it} \)

\( \sigma^2_{a} \)

\( \alpha_{0} \)

\( \alpha_{1} \)

\( \alpha_{2} \)

\( m_{it}^{a} \)

\( a_{i} \)

\( a_{it} \)

\( \sigma^2_{a} \)

\( \alpha_{0} \)

\( \alpha_{1} \)
patterns will converge to some \textit{post-information} level. \( \frac{3}{2} \), the variance in the unobserved component of consumers' known tastes for Yoplait 150, is identified by unobserved heterogeneity in consumers' \textit{pre-information} (pre-first purchase) behavior. On the other hand, \( \frac{2}{2} \), the variance of the unknown taste \( \pm \) across the population, is identified by comparing the variance of \textit{pre-information} heterogeneity to the variance of \textit{post-information} heterogeneity. \( \pm \) the mean experience utility of Yoplait 150, is assessed by a comparison of the means of these two distributions, i.e. whether \textit{post-information}, consumers (on average) purchase Yoplait 150 more or less than \textit{pre-information} (net of experimentation behavior due to dynamic optimization). \( \frac{2}{2} \), the per-period variance in experience utility, is identified by the number of consumption experiences it takes for consumers to learn \( \pm \), i.e. how many consumption experiences it takes for purchasing patterns to converge to the \textit{post-information} level. If, for example, purchase patterns change after initial purchases, but not thereafter, it is indicative that \( \frac{2}{2} = 0 \), i.e. a one-consumption learning process. The advertising-related coefficients, \( \mu_1 \) and \( \mu_2 \), are identified by the effects of advertising exposures on inexperienced and experienced consumers (both in an absolute sense \( \mu_x \)) and relatively \( \mu_1 \).

4.3. Estimation

Moving to estimation, the primary complication is consumer heterogeneity and the resulting number of econometric unobservables. Besides the per-period logit errors, we do not observe a consumer's \( \pm \), his random component of \( \xi_t \), and his realizations of experience utility \( \pm_{t+1} \) at each purchase occasion. Recall that these unobservables are assumed mutually uncorrelated except for the fact that experience utility realizations are distributed around \( \pm \). In addition, these unobservables are assumed independent of our observables \( y_t; a_{xt}; p_{xt}; sc_{xt}, \) and \( p_{yt}^{th} \). Because of the persistent unobservables and

\footnote{The last “learning process” parameter, \( \sigma^2 \), is identified through its appearance in the prior variance matrix. Adjusting this affects 1) experimentation behavior in the dynamic model, and 2) the shape of learning (how posteriors evolve over time).}

\footnote{Perhaps the most likely violation of this assumption would be due to the endogeneity of supermarket choice and shopping trip timing. We cannot have consumers getting high \( \epsilon_{xt} \) draws and searching out low Yoplait 150 prices. Hopefully Yogurt is a small enough component of consumers’ purchases to prevent significant such behavior. Another potential endogeneity problem arises if firms are able to focus advertising towards consumers who like Yoplait 150 more than our observables predict (i.e. \( a_{xt} \) is correlated with the random component of \( \lambda_t \) or \( \delta_t \)). Ackerberg (2001) did not find statistical support for this possibility.}
the dependence of purchase probabilities on lagged endogenous variables (through posteriors), we use Simulated Maximum Likelihood, integrating the persistent unobservables over the entire sequence of a consumer’s choices to derive the probability of that consumer’s observed data. This results in the following likelihood function (for consumer i):

\[ L_i(\mu) = \Pr \left( c_{it} = c(m_{it}(\pm; a_{it}; c_{it}^{-1}; \mu); z_{it}; S_{it}; \ldots; z_{it}^{-1}; \mu) \right) \]

\[ = \Pr \sum_{i=1}^{T_i} \prod_{t=1}^{T_i} \Pr \left( c_{it} = c(m_{it}(\pm; a_{it}; c_{it}^{-1}; \mu); z_{it}; S_{it}; \ldots; z_{it}^{-1}; \mu) \right) \]

\[ = R Q_{T_i} \prod_{t=1}^{T_i} \Pr \left( c_{it} = c(m_{it}(\pm; a_{it}; c_{it}^{-1}; \mu); z_{it}; S_{it}; \ldots; z_{it}^{-1}; \mu) \right) \]

where \( c_{it} \) is the consumer’s observed choice in period \( t \), \( c(\phi) \) is the model’s predicted choice, \( z_{it} = (p_{it}; p_{it}^{th}; s_{it}) \), \( T_i \) is the total number of shopping trips of consumer \( i \), and superscripts indicate histories of a variable through that point (e.g. \( a_{it}^{'} = a_{i1}; \ldots; a_{it} \)). The \( \Pr[\phi] \) in the last line is the probability that the period \( t \) logit errors (\( \epsilon_{it} \)) are such that the model’s predicted choice equals our observed choice, conditional on \( z_{it} \), past choices \( c_{it} \), and past realized \( \epsilon_{it+1} \)’s. The predicted choice function \( c(\phi) \) is defined by either myopic utility maximization or the optimal policy function generated by the dynamic programming problem. Under the i.i.d. logit assumption on the \( \epsilon_{it} \)'s, the last \( \Pr[\phi] \) has a closed form solution in both the myopic and fully dynamic cases (Rust (1987)).

4.4. Dynamic Programming Solution

In estimating the full dynamic model, we must solve the consumer’s dynamic programming problem to obtain \( c(\phi) \). As this solution depends on the majority of the model’s parameters, it needs to be embedded into the routine used to maximize the likelihood function. Our utility specification and the

\[ ^{28} \text{With the exception of the logit errors, the integrals generated by these unobservables are not analytically computable and we rely on either simulation or discrete approximations to evaluate them. As is well known (e.g. Keane (1994)), simulation of these integrals combined with ML estimation results in inconsistent estimates for a finite number of simulation draws.} \]
assumed learning process generate the Bellman's equation:

\[ V(s_{it}; \mu) = \max \mathbb{E} \left[ U_{it} + \bar{V}(s_{it+1}; \mu) | s_{it}; c_{it} = 1 \right] + U_{it} + \mathbb{E} \left[ V(s_{it+1}; \mu) | s_{it}; c_{it} = 2 \right] \]

where the state space \( s_{it} = (\_i; m_{it}; \bar{s}_{it}; p_{it}; sc_{it}; p^{th}_{it}; \bar{\omega}_{it}; \bar{\omega}_{it+1}) \). Although this state space appears to be quite large, there are a number of simplifications and assumptions that we use to significantly reduce the dimensionality of the problem and allow for relatively quick numeric solution. First, conditional on \( \bar{s}_{0t} \) and \( \bar{\omega} \), \( K_{it} \) (the number of purchases up to \( t \)) and \( TV_{it} \) (total hours of television watched up to \( t \)) are sufficient to define the posterior variance matrix \( \bar{s}_{it} \). Therefore, our assumption that \( \bar{s}_{0t} \) and \( \bar{\omega} \) are constant across consumers\(^{29} \) allows us to replace \( \bar{s}_{it} \) in the state space by \( K_{it} \) and \( TV_{it} \): Second, because \( \bar{\omega}_{t+1} \) enters the utility function linearly, \( \_i \) can be merged into the learning process (so our consumer Bayesian updates on the sum \( \_i + \_i \)). Third, state variables whose realizations only affect current utility need not be solved for explicitly as state variables (Rust (1987), Keane and Wolpin (1994)). This removes the i.i.d. \( \bar{\omega}'s \) from the effective state space, and as we assume that consumers perceive \( p_{it}; sc_{it} \) and \( p^{th}_{it} \) to be non-serially correlated\(^{30} \), we end up with a four dimensional problem where \( s_{it} \) the effective state space equals (\( m_{it}; K_{it}; TV_{it} \):

A second major simplification results from the existence of an analytic solution for the expected value of the maximum of logit errors. As a result, the expectation over future \( \bar{\omega}'s \) in the Bellman equation can be computed analytically (Rust(1987)). This, along with the assumption of discrete perceived distributions of future \( p_{it}; sc_{it}; and p^{th}_{it} \), implies that to compute the expectations in the value function we need only numerically integrate over the distribution of next period's posterior means:

\(^{29}\)For \( \Phi \) to be constant, we assume that all consumers anticipate watching the same amount of television between the current and next shopping (the sample mean). This assumption only affects perceptions of the future - in the actual learning process, \( \Phi \) depends on the amount of television watched between \( t - 1 \) and \( t \). A similar assumption is that consumers use the same per-shopping trip discount factor to weight the future. Again, this only affects perceptions of the future. These assumptions were made for computational reasons - relaxing either of would at the very least add an extra state variable to the model.

\(^{30}\)We assume a 4-point (estimated from the data) perceived distribution of future \( p_{it} \), and a degenerate distribution of \( sc_{it} \) and \( p^{th}_{it} \) (i.e. consumers expect that next shopping trip, \( sc_{it} \) and \( p^{th}_{it} \) will be at their respective means). The assumption that the \( p_{it} \) are iid is easily rejected by the data, and is only adopted for computational reasons. Hopefully the effect on estimation results is small, as this distribution only enters into expectations of the future in the dynamic program (and doesn’t enter the myopic estimation results at all).
Because the elements of \( s_t \) are either continuous variables \( (m_{it}) \) or take on a large number of discrete values \( (K_{it} \text{ and } TV_{it}) \), the state space must be discretized in order to apply the Method of Successive Approximations and numerically solve the above dynamic programming problem. We do not discretize the entire dynamic problem, but choose points at which we will solve for the (approximate) value function. In the following estimation results we have discretized the state space into 10 to 20 points in each dimension. Because the numerical integration mentioned above is only two dimensional we chose to use quadrature rather than Monte-Carlo. Since the quadrature points generate future states that are not our discretized ones, we use linear interpolation to evaluate the value function at these states.

4.5. Results

Table 4.1 presents maximum likelihood estimates of the above model. In initial runs we had trouble obtaining reasonable estimates of \( \frac{\sigma^2}{\sigma} \), the per-period variance in experience utility. Our estimates were unreasonably high, indicating that consumers were learning (through consumption) about the unobservable characteristics of Yoplait 150 very slowly. As this parameter is identified by changes in consumer purchasing patterns after consumption experiences, it is likely that it picks up other unobservables that cause such changes (e.g. learning about other brands or products). As a result, the majority of our estimates assume a one-period learning process \( \frac{\sigma^2}{\sigma} = 0 \). This assumption has the added benefit of greatly reducing both the computational burden of the dynamic programming problem and likelihood evaluation. We have capitalized on this computational reduction to increase the precision and accuracy of our discretization and integral evaluation over what would have otherwise been possible\(^{31}\).

The first two columns of Table 4.1 contain results under the myopic assumption on behavior,

\(^{31}\)With this assumption, the state variable “number of previous purchases” becomes a simple indicator variable whether the consumer has ever bought Yoplait 150. For likelihood evaluation, we now only need to numerically evaluate a two dimensional integral for each household (rather than \( 2 + K_{iT} \)). Then, with the following discretization of the state space \( m_{it} - 20 \text{ points}, \ m_{it} - 10 \text{ points}, \ K_{it} - 2 \text{ points (either purchased or never purchased), } TV_{it} - 10 \text{ points}), 49 \text{ draws on the Bellman's equation integral (for a total of 196,000 3-dimensional interpolations per contraction iteration), and each of the distributions of } \delta_i \text{ and the random component of } \lambda_i \text{ discretized into a 26 point normal for likelihood evaluation (for a total of 26^2 draws), one function evaluation with a discount rate of } .98 \text{ on an UltraSparc 167Mhz processor in highly optimized C code takes about 7 minutes (~5.5 min. for the contraction mapping to converge, 1.5 min. for likelihood evaluation).}
with and without a time trend\textsuperscript{32} on the outside alternative. The last three columns are results from estimation of the full dynamic model. In all ve models, the price, other price, and store coupon coefficients are very similar, significant and the hypothesized sign. The price coefficients generate price elasticities of demand for Yoplait 150 (over the entire time frame of the sample) of approximately 3.3. The estimates and significance of $\beta$ indicate that there is significant unexplained heterogeneity in consumers' initial valuations of Yoplait 150. The very large significance of the estimates of $\beta$ strongly support the existence of imperfect information and learning, indicating that consumers' have heterogeneous components of utility that are not realized until after their rst consumption experience with Yoplait 150.

Moving from the myopic to the fully dynamic models corresponds to allowing the consumers' discount factors to differ from zero. In the model without the time trend, this results in a significant increase in likelihood and an estimated discount factor of .981. As this is a discount factor for the time between shopping trips, which averages only a little more than a week, this estimate is low, though not necessarily unreasonable. It may be capturing consumer uncertainty on how long Yoplait 150 might remain on the market or the possibility of newer, better Yogurts being introduced\textsuperscript{33}.

Of particular interest in comparing the myopic and fully dynamic results are the estimates of $\beta$, the mean experience quality of Yoplait 150. While the negative estimates in the myopic models suggest that consumers (on average) liked Yoplait 150 less than expected, the positive dynamic estimates indicate that consumers (on average) were pleasantly surprised by the experience quality of the brand. Although these coefficients are not generally significant, this points out a possible bias in the myopic assumption. Experimentation behavior generated by a true dynamic decision process is likely interpreted in a myopic model as overpredictions of experience utility by consumers.

The estimated coefficients pertaining to advertising strongly support the hypothesis that adver-

\textsuperscript{32}This time trend, for example, might capture effects of new entrants into the yogurt market over the period. Adding a linear time trend to the dynamic problem does not increase computation as the trend can be merged into posterior means. With linearity, expectations of future means look the same from any point in time.

\textsuperscript{33}In the dynamic models including a time trend, reasonable values (above .9) of the discount factor result in worse likelihoods than the corresponding myopic model. This is likely the result of the positive time trend on the outside good generating behavior similar to the experimental behavior generated by a higher discount factor. We therefore fix the discount factor at its point estimate from the dynamic model without the time trend.
tising affects consumers mainly through the informational structure of the model. Our \textit{prestige} coefficient on advertising directly in the utility function, $\mu$, is actually significantly negative in the models without a time trend, though this may be due to an upward time trend in posterior advertising means. In the models including a time trend, the coefficients are virtually zero and insignificant.

For interpretation purposes, we report estimates of $1/2$ the correlation coefficient associated with $\delta_0$; rather than $-1$. All estimates of $1/2$ are significantly positive, suggesting that advertising is providing consumers with information. The parameters at the bottom of the table provide an indicator of the magnitude of this informative effect. These indicate the percent increase in purchase probability of a representative inexperienced consumer after a doubling of advertising exposures for 1, 5, and 20 week periods. These informative advertising \textit{elasticities} are all significantly positive, suggesting that advertising is providing consumers with information. The parameters at the bottom of the table provide an indicator of the magnitude of this informative effect. These indicate the percent increase in purchase probability of a representative inexperienced consumer after a doubling of advertising exposures for 1, 5, and 20 week periods. These informative advertising \textit{elasticities} are all significantly positive, suggesting that advertising is providing consumers with information.

As a check of the goodness of fit in our model, Table 4.1b compares conditional choice probabilities predicted by the model (the last column of Table 4.1) to conditional choice probabilities in the data. We condition on number of past purchases, from 0 to 10. Casually, the model simulations appear to match the data reasonably well - statistically, only 2 of the 11 data probabilities are significantly different from the model probabilities.

Table 4.2 examines three perturbations of the myopic model (these perturbations make the dynamic model infeasible to estimate). In the first column we allow a more flexible persuasive effect of advertising, including dummies for $m_{it}$ lying in different regions. The estimates are imprecise. We also tried using non-linear functional forms for $m_{it}$, but did not find much. The second column includes a random coefficient on persuasive advertising, allowing heterogeneity in prestige or image effects across

\footnote{That is, doubling Yoplait's advertising level over the entire time frame of the data results in a 15% increase in sales. Along with the estimated price elasticity, this implies an advertising to sales ratio of 4.5% in a static, single-product firm, advertising and price-setting model (where the ad to sales ratio equals $\frac{A}{P}$). Though these seem to be reasonable results (According to Advertising Age, in 1988 total Yoplait advertising expenditures were about 7% of total sales), this static firm side model is obviously inefficient.}
the population. Although the estimated standard deviation of the random coefficient is economically large it is insignificant even when simulation error is neglected. The mean prestige effect stays at essentially zero. We hesitate to make any strong conclusions about these results because the random coefficient may be picking up measurement error in our advertising variable. The third column relaxes our assumption of a one-period learning process, \( \gamma_2 = 1 \) (This results in a path of posterior experience quality variances (in # of prior purchases) of 2.882, .742, .426, .299, .230, .187,....). Again, we obtain an insignificant prestige effect and a positive, significant informative effect. In summary, our results suggest that: (1) there were little if no Becker-like prestige effects generated by these advertisements, and (2) these advertisements provided consumers with some type of product information, not that this was necessarily signaling information.

5. Welfare Analysis

We now move to a welfare analysis of the above results, examining the social welfare consequences of advertising in this market. In contrast to the above, this analysis is highly conditional on the assumption that the informative effect of advertising we have found is in fact pure signaling information.

There is a relatively clear cost-benefit trade-off with our signaling effect of advertising: costs the cost of resources devoted to advertising, benefits the information conveyed to consumers by the advertising (we ignore potential benefits (or costs) of advertising outside this market, in particular its subsidization of media.) On the other hand, for possible prestige effects, there is a serious question as to how to measure potential benefits of advertising. Because utility is a latent variable, a positive prestige effect only indicates that advertising increases the utility of consuming Yoplait 150 relative to other yogurts. We cannot distinguish whether it adds to the utility obtained from consuming Yoplait 150 or subtracts (i.e. dis-prestige) from the utility derived from consuming the outside alternative.

\[^{35}\]Because there are 3 unobservables to integrate out over in this model, we switch from discretizing the integrals to using Monte-Carlo with crude importance sampling. We also use this technique in the next set estimates, where there are \((2+k_{IT})\) dimensional integrals to evaluate. As a result of this, the likelihood values in the last two columns are not directly comparable with the previous likelihoods.

\[^{36}\]It is also dependent on numerous and likely simplistic firm behavioral assumptions and equilibrium extrapolations that will follow. As a result we do not take this to be the final conclusion on the welfare effects of advertising in this market. However, we do feel is that it is an interesting exploration of the structural estimates and illuminate interesting (and typically ignored) possibilities concerning the welfare effects of informative advertising.
Thus our empirical result that there is no prestige effect of advertising in this market is facilitating here.

Our first step is to combine the final dynamic estimates of the demand for Yoplait 150 (the last column of Table 4.1) with profit maximizing first order conditions of Yoplait to back-out the production cost of Yoplait 150 and costs of advertising. Unfortunately, the first-order conditions which we would like to have, those arising from Yoplait’s dynamic price and advertising setting problem, are not feasibly obtainable within the context of the fairly complicated consumer side model presented here. Even ignoring other products, the firm’s state space for such a dynamic model would need to contain the joint distribution of consumer tastes and experience. This is far too complicated for the present work and we proceed using a major simplification of producer behavior: one in which the firm sets one price and one advertising level in order to maximize profits in some introductory period.

Another problem is that we do not observe a single price, but rather a distribution of prices. Again in order to simplify things, we assume that the firm chooses a mean price with the price distribution around that mean price fixed (as observed in our sample - perhaps due to retailer behavior).

The above assumptions result in Yoplait setting its mean price $p$ and advertising intensity $a$ to maximize total profits over the introductory period:

$$\pi(p, a) = TR(p, a) - q(p, a) mc - c_a$$

where $q(p, a)$ and $TR(p, a)$ are total sales and total sales revenue, $mc$ is the assumed constant marginal cost of a unit of Yoplait 150, and $c_a$ is the cost per unit of advertising intensity. $TR(p, a)$ and $q(p, a)$ are quantities which we can simulate using our estimated structural model. This involves drawing consumers from the estimated distribution of consumer heterogeneity, simulating prices and advertising exposures, and computing information paths and optimal purchase decisions through the time frame of the model. Note that $TR$ cannot be decomposed (into $p \cdot q$) because of the distribution

\textsuperscript{37}We take this introductory period to be the length of our data. One possible justification for this could be that product qualities are somehow revealed to consumers after the introductory period, essentially removing any dynamic effects of current price and advertising setting on profits after this period.
of prices. Differentiating with respect to both $p$ and $a$ and manipulating the two F.O.C.'s gives:

\[
mc = \frac{\partial TR(p, a)}{\partial p} \frac{\partial \bar{q}(p, a)}{\partial p}
\]

\[
c_a = \frac{\partial TR(p, a)}{\partial a} \frac{\partial \bar{q}(p, a)}{\partial a} mc
\]

Using our estimates we simulate these four derivatives and solve these equations, obtaining a marginal
cost of Yoplait 150 of $0.422 (p = $0.653) and a cost per unit of advertising intensity of $3732^{38}.

With these costs in hand, we can analyze the welfare effects of a ban on advertising. We first
eliminate the covariance term in consumers' priors (i.e. set $\bar{\theta}_0 = \bar{\theta}_1 = 0$). This constitutes rational
consumer behavior under the new regime, as the resulting advertising intensity of 0 should not tell
consumers anything about Yoplait 150's experience quality. Because of the dynamics, this loss of
advertising information also results in consumers who are more inclined to "experiment" with Yoplait
150 in order to learn about its experience characteristics. With $\bar{\theta}_0 = \bar{\theta}_1 = 0$ we obtain a new demand
system ($TR(\phi)$ and $q(\phi)$ functions), which is used to numerically "nd the new $p$ which solves the above
`rst order conditions (unfortunately, we cannot optimally adjust competitors' prices because we have
no model of demand for competing products). We "nd a new pro`t maximizing mean price of Yoplait
150 of $0.646. The near one cent reduction in optimal price results from Yoplait 150's better than
average mean experience characteristics ($\pm :899 > 0)$. The advertising ban prevents Yoplait from
signaling this through advertising and reduces their ability to price this quality.

The first two columns of Table 5.1 exhibit some welfare measures of the two equilibria. Both `rm
pro`ts and compensating variation (CV) increase under the advertising ban. Yoplait does make less
variable pro`ts, but this is overcome by saved advertising expenditures. The change in CV measures
two effects on consumers: 1) the loss of the signaling information contained in advertising and 2) the
equilibrium change in price. To separate the two effects, the third column considers the case when
Yoplait cannot adjust mean price. The change in CV from the `rst column to here measures the pure

\footnote{This advertising cost is given our simulated dataset of 1.2 million consumers. Given we assume constant marginal
costs of production, we could multiply by the U.S. population/1.2 million to get national numbers. (Obviously, one needs
to assume that Springfield and Sioux Falls are representative of the entire U.S. population here).}
value of the advertising information. Though this information value is positive, its loss is more than outweighed by Yoplait 150's price cut. This analysis suggests that the costs of advertising Yoplait 150 far exceeded its informational benefits.

On the other hand, this is only part of the story. Yoplait 150 is just one product introduction with one particular ±. The linear signaling equilibrium we have estimated is an equilibrium over many possible new products with many different experience qualities ±. The benefits of signaling information should differ for different ±s. Intuitively, the benefit should be greatest for products with experience qualities far away from consumers' initial priors on ± as these are the products for which our advertising can provide the "most" information. Thus, to completely assess the welfare consequences of this signaling equilibrium, what we really want to do is integrate welfare benefits over the distribution of all possible product introductions (i.e. all possible experience qualities). Assuming consumers are "correct" (in a probabilistic sense) in their priors, this distribution is actually part of our consumer model - it is the consumers' initial prior on ± (which we have assumed normal, normalized its mean to 0, and estimated the standard deviation 3/4 = 0.593).

Unfortunately, we have no direct data on these other "products" with different ±s. However, our estimated model allows us to compute demand for alternate experience qualities (TR(p; a; ±), q(p; a; ±)). In addition, although we do not know marginal costs and optimal mean prices for alternative experience qualities (mc(±) and p(±)), our estimated signaling equilibrium equation does tell us how much they should optimally advertise (a(±) = −0 + −1 ±). Therefore, as we have two first order conditions and just two unknowns (mc and p), we can numerically solve out these unknowns for each ±. Knowing mc(±); we can then consider an advertising ban, inverting out prices under the ban. Figure 5.1 plots the results: p(±); mc(±); and p\textsuperscript{noad}(±) - all increase in experience quality\textsuperscript{39}. Of particular note is the result that under the ban, prices become more equalized. Due to the lessened information, higher

\textsuperscript{39}The fact that prices vary over experience quality raises the question of why consumers can't simply infer quality from price. One possibility is that there is too much variation in the price distribution (although there is variance in advertising also). Another is that, as in Milgrom and Roberts, firms in equilibrium need to set both prices and advertising levels appropriately for a credible signal. A more logistical problem is that our linear advertising equilibrium equation indicates low δ\textsubscript{j}'s (those more than 1.54 standard deviations below the mean) should advertise negative amounts. In simulating welfare for these qualities, we assume that no money is spent on advertising but that consumers get the correct signals anyway, again likely biasing our results slightly in favor of advertising.
quality products are less able to price their quality while lower quality products can extract more "dis-information" rents.

Figure 5.2 plots CV under the two regimes, as well as for the case where the rms do not change price under the ban (again to separate out information effects from price effects). As expected, the further experience quality is away from their prior mean the more consumers benefit from the advertising information (Figure 5.2b). In fact, for experience qualities close to 0, the effect of advertising information on compensating variation is negative because the noise in the advertising signal moves consumers away from initially close-to-correct priors\(^1\). The price increases coming with the ban at the low end of the spectrum accentuate the loss in CV, while the price decreases at the high end more than compensate for it (Though it appears that above ±1.4 they may not).

Figures 5.3 - 5.6 indicate sales, revenues, costs, and profits under the two regimes. Most notable is the fact that profits go up under the advertising ban over the entire range of qualities. Again this suggests that this equilibrium between consumers and rms is such that rms are hurt by the ability to advertise, at least when we only account for profits in this introductory period. Also of note is the fact that profits decrease in experience quality. Again these are only introductory period profits, and there may be compensating positive returns to experience quality after the introductory period, but this result is somewhat unappealing\(^1\). More appealing is the fact that the advertising ban slightly accentuates the slope of the profit curve. This suggests that the ability to advertise increases incentives (or at least decreases disincentives!) to invest in experience quality in a model where such choices were endogenized. Figure 5.7 plots total surplus under the two regimes; only in the very negative range do consumer losses outweigh profit gains. The approximate value of the integral of the welfare gain over the ± distribution divided by total revenues (similarly weighted) implies a welfare

\(^{1}\)It appears that the lack of symmetry of the "value of information" function (the minimum being at experience quality .2) arises from a somewhat inconsistent treatment of the discount factor. In adding our welfare measures over periods we use discount factor above the estimated .98 (otherwise things die out very quickly). This means that consumers are actually not behaving to exactly maximize our CV measure. This lower discount factor implies less experimentation behavior, and therefore from our social planner standpoint it is better to fool the consumers into a bit more experimentation.

\(^{1}\)Generating this result is the fact that lower experience quality products' lower prices generate many more first-time purchases and many more idiosyncratic taste draws from a fairly high variance distribution. We suspect that this slope might disappear or change sign if one allowed firms some flexibility in changing prices over time.
gain to the ban of slightly more than 4% of industry sales, suggesting that if this is in fact signaling information, it is not providing the information very efficiently.

6. Conclusions

In summary, we feel our structural estimation results are thought provoking. We present a model in which we explicitly include two effects of advertising: an informative effect which enters the information structure of our dynamic consumer learning model, and a prestige or image effect entering directly into consumers' utility functions. Structural estimation of this model finds a large, significant, and robust informative effect of advertising and an insignificant prestige effect, suggesting that these Yoplait 150 television advertisements affected consumers primarily through the provision of information, not through prestige or image effects. These results support the conclusions of Ackerberg (2001) and strengthen them by explicitly allowing and controlling for experience characteristics and consumer learning. We feel that together the two approaches provide a broad framework within which one can analyze effects of advertising for other products, given the appropriate data. Of particular interest might be comparing estimates across different types of products, seeing if one can find the existence of prestige or image effects.

An important next step is to ascertain what the implications are of such findings on the functioning of markets. We feel that knowledge of how advertising affects or potentially affects a market should be an important consideration in policy decisions with respect to that market. We take a brief stab at such questions in our welfare examination, albeit in a somewhat unsatisfying way as the analysis rests on some very simple and strong assumptions on firm behavior and our particular modeling of informative advertising. The lack of a realistic firm side model is also problematic because it creates an inability to convincingly consider dynamic decisions such as entry and innovation. These are two very interesting and policy relevant variables that are likely to depend on the way or ways in which advertising works in a market.

These deficiencies point to further research. One direction is moving to consumer levels models incorporating multiple informative effects of advertising. As suggested in Ackerberg (2001), different
informative effects can potentially be distinguished with the proper data. Perhaps more challenging is developing realistic empirical models of firm behavior in markets with imperfect information and advertising. Such models would need to be dynamic, as decisions such as price have dynamic effects though their effects on consumer information. We also would want such models to endogenize entry and innovation. Unfortunately, it is likely not feasible to embed a consumer demand model as rich as the above into a dynamic model of firm behavior. Therefore, the challenge is to develop a demand side rich enough to accommodate such decisions and effects but parsimonious enough to be able to solve and estimate.
Table 3.1: Summary Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Market 1</th>
<th>Market 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Households</td>
<td>950</td>
<td>825</td>
</tr>
<tr>
<td>Total Shopping Trips</td>
<td>67051</td>
<td>54308</td>
</tr>
<tr>
<td>Average Shopping Trips per Household</td>
<td>70.58</td>
<td>65.82</td>
</tr>
<tr>
<td></td>
<td>(33.39)</td>
<td>(31.82)</td>
</tr>
<tr>
<td>Average Price of Yoplait 150 (Cents)</td>
<td>0.645</td>
<td>0.663</td>
</tr>
<tr>
<td></td>
<td>(0.060)</td>
<td>(0.079)</td>
</tr>
<tr>
<td>Shopping Trips with Yoplait 150 Purchase</td>
<td>302</td>
<td>650</td>
</tr>
<tr>
<td>Manufacturers Coupons Redeemed for Yoplait 150</td>
<td>16</td>
<td>238</td>
</tr>
<tr>
<td>Shopping Trips with Other Yogurt Purchase</td>
<td>5432</td>
<td>3863</td>
</tr>
<tr>
<td>Households Trying Yoplait 150</td>
<td>123</td>
<td>184</td>
</tr>
<tr>
<td>Households Trying Other Yogurts</td>
<td>648</td>
<td>512</td>
</tr>
<tr>
<td>Commercial Exposures</td>
<td>12918</td>
<td>12563</td>
</tr>
<tr>
<td>Commercial Exposures per Household</td>
<td>13.60</td>
<td>15.22</td>
</tr>
<tr>
<td></td>
<td>(10.81)</td>
<td>(9.96)</td>
</tr>
<tr>
<td>Advertising Share of Yoplait 150</td>
<td>0.35</td>
<td>0.37</td>
</tr>
<tr>
<td>Market Share of Yoplait 150</td>
<td>0.05</td>
<td>0.14</td>
</tr>
</tbody>
</table>

Note: Standard Errors in parentheses where applicable
### Table 4.1: Myopic and Full Dynamic Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Myopic Model</th>
<th>Myopic Model w/ Time Trend</th>
<th>Dynamic Model</th>
<th>Dynamic Model w/ Time Trend</th>
<th>No Prestige Advertising</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1$ - Price</td>
<td>-5.26140</td>
<td>-5.54170</td>
<td>-5.29900</td>
<td>-5.49350</td>
<td>-5.48930</td>
</tr>
<tr>
<td></td>
<td>(0.31620)</td>
<td>(0.32557)</td>
<td>(0.31454)</td>
<td>(0.33230)</td>
<td>(0.32980)</td>
</tr>
<tr>
<td>$\theta_2$ - Store Coupon</td>
<td>3.11930</td>
<td>3.11540</td>
<td>3.04460</td>
<td>3.11030</td>
<td>3.09590</td>
</tr>
<tr>
<td></td>
<td>(0.87961)</td>
<td>(0.80984)</td>
<td>(0.83679)</td>
<td>(0.81199)</td>
<td>(0.80927)</td>
</tr>
<tr>
<td>$\theta_3$ - Prestige Advertising</td>
<td>0.10537</td>
<td>0.00281</td>
<td>-0.13855</td>
<td>-0.02469</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>(0.03751)</td>
<td>(0.03886)</td>
<td>(0.04117)</td>
<td>(0.04415)</td>
<td></td>
</tr>
<tr>
<td>$\theta_4$ - Competitor’s Price</td>
<td>-0.74010</td>
<td>-0.69667</td>
<td>-0.77477</td>
<td>-0.69536</td>
<td>-0.70704</td>
</tr>
<tr>
<td></td>
<td>(0.22169)</td>
<td>(0.22243)</td>
<td>(0.22154)</td>
<td>(0.22136)</td>
<td>(0.21827)</td>
</tr>
<tr>
<td>$\theta_5$ - Time Trend on Outside Alternative</td>
<td>0</td>
<td>1.16370</td>
<td>0</td>
<td>0.94299</td>
<td>0.98856</td>
</tr>
<tr>
<td></td>
<td>(0.17081)</td>
<td>(0.16134)</td>
<td>(0.13694)</td>
<td>(0.13694)</td>
<td></td>
</tr>
<tr>
<td>$\sigma_i$ - Variance of $\delta_i$ around $\delta$</td>
<td>1.76690</td>
<td>1.77900</td>
<td>1.86030</td>
<td>1.83750</td>
<td>1.81620</td>
</tr>
<tr>
<td></td>
<td>(0.13582)</td>
<td>(0.13985)</td>
<td>(0.13924)</td>
<td>(0.13319)</td>
<td>(0.13261)</td>
</tr>
<tr>
<td>$\sigma_i$ - Consumer’s perceived variance of $\delta$</td>
<td>0.2348</td>
<td>1.65380</td>
<td>1.88830</td>
<td>0.64559</td>
<td>0.59278</td>
</tr>
<tr>
<td></td>
<td>(0.23840)</td>
<td>(1.57685)</td>
<td>(1.33743)</td>
<td>(0.32074)</td>
<td>(0.27870)</td>
</tr>
<tr>
<td>$\rho$ - Correlation Coefficient of $x_0$ - Informative Advertising</td>
<td>0.36563</td>
<td>0.67287</td>
<td>0.14273</td>
<td>0.32583</td>
<td>0.12317</td>
</tr>
<tr>
<td></td>
<td>(0.12444)</td>
<td>(0.36086)</td>
<td>(0.09219)</td>
<td>(0.04054)</td>
<td>(0.03619)</td>
</tr>
<tr>
<td>$\delta$ - Mean Experience Quality of Yoplait 150</td>
<td>-2.41900</td>
<td>-0.71716</td>
<td>0.67338</td>
<td>0.80500</td>
<td>0.89878</td>
</tr>
<tr>
<td></td>
<td>(0.72998)</td>
<td>(0.99348)</td>
<td>(0.78308)</td>
<td>(0.28409)</td>
<td>(0.27600)</td>
</tr>
<tr>
<td>Discount Factor</td>
<td>0</td>
<td>0.98139</td>
<td>0.98</td>
<td>0.98</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.01885)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_1$ - Constant</td>
<td>-4.45980</td>
<td>-3.41110</td>
<td>-4.88810</td>
<td>-4.51520</td>
<td>-4.60370</td>
</tr>
<tr>
<td></td>
<td>(0.97585)</td>
<td>(1.08139)</td>
<td>(0.53794)</td>
<td>(0.46500)</td>
<td>(0.46537)</td>
</tr>
<tr>
<td>$\lambda_2$ - Market Dummy</td>
<td>1.65010</td>
<td>1.49190</td>
<td>1.22720</td>
<td>1.25530</td>
<td>1.27340</td>
</tr>
<tr>
<td></td>
<td>(0.19015)</td>
<td>(0.17476)</td>
<td>(0.23333)</td>
<td>(0.17639)</td>
<td>(0.17298)</td>
</tr>
<tr>
<td>$\lambda_3$ - Income</td>
<td>0.08351</td>
<td>0.07467</td>
<td>0.05884</td>
<td>0.06475</td>
<td>0.05995</td>
</tr>
<tr>
<td></td>
<td>(0.03342)</td>
<td>(0.03114)</td>
<td>(0.02958)</td>
<td>(0.02679)</td>
<td>(0.02658)</td>
</tr>
<tr>
<td>$\lambda_4$ - Family Size</td>
<td>-0.07470</td>
<td>-0.06929</td>
<td>-0.08792</td>
<td>-0.06254</td>
<td>-0.02484</td>
</tr>
<tr>
<td></td>
<td>(0.08044)</td>
<td>(0.07055)</td>
<td>(0.06345)</td>
<td>(0.06048)</td>
<td>(0.06061)</td>
</tr>
<tr>
<td>$\lambda_5$ - Presample</td>
<td>0.01494</td>
<td>0.01380</td>
<td>0.01185</td>
<td>0.01148</td>
<td>0.01087</td>
</tr>
<tr>
<td>Yogurt Purchases</td>
<td>(0.01485)</td>
<td>(0.01326)</td>
<td>(0.01124)</td>
<td>(0.01113)</td>
<td>(0.01113)</td>
</tr>
<tr>
<td>$\lambda_6$ - Presample</td>
<td>-0.00014</td>
<td>-0.00012</td>
<td>-0.00011</td>
<td>-0.00010</td>
<td>-0.00010</td>
</tr>
<tr>
<td>Yogurt Purchases$^2$</td>
<td>(0.00002)</td>
<td>(0.00002)</td>
<td>(0.00002)</td>
<td>(0.00002)</td>
<td>(0.00002)</td>
</tr>
<tr>
<td>$\lambda_7$ - Presample</td>
<td>0.04636</td>
<td>0.04687</td>
<td>0.04216</td>
<td>0.04108</td>
<td>0.04219</td>
</tr>
<tr>
<td>Yogurt Purchases</td>
<td>(0.01583)</td>
<td>(0.01410)</td>
<td>(0.01319)</td>
<td>(0.01195)</td>
<td>(0.01195)</td>
</tr>
<tr>
<td>$\lambda_8$ - Presample</td>
<td>0.04221</td>
<td>0.03549</td>
<td>0.03272</td>
<td>0.03090</td>
<td>0.03136</td>
</tr>
<tr>
<td>Lowfat Purchases</td>
<td>(0.01667)</td>
<td>(0.01487)</td>
<td>(0.01322)</td>
<td>(0.01252)</td>
<td>(0.01256)</td>
</tr>
<tr>
<td>$\sigma_\lambda$</td>
<td>2.13160</td>
<td>1.76610</td>
<td>1.62990</td>
<td>1.51500</td>
<td>1.51040</td>
</tr>
<tr>
<td></td>
<td>(0.17652)</td>
<td>(0.16045)</td>
<td>(0.28756)</td>
<td>(0.22596)</td>
<td>(0.21967)</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-395.83624</td>
<td>-394.3477</td>
<td>-395.6524</td>
<td>-394.6655</td>
<td>-394.6053</td>
</tr>
</tbody>
</table>

Note: Below parameters not estimated, standard errors obtained by delta method

- Informative Advertising
  - Effect - 1 Week: 0.20384
  - Effect - 5 Weeks: 0.38684
  - Effect - 20 Weeks: 0.69631

- Informative Advertising
  - (0.11962) 0.17159 0.08390 0.07528 (0.06206)
  - (0.17380) 0.16671 0.18483 0.14683 (0.12727)
  - (0.23173) 0.18467 0.32990 0.17113 (0.14201)

- [X]_11^2 = (∑i=1^3σi)^2/2 = 1.99367 2.43135 2.65073 1.94761 1.91048

(0.16888) 0.10819 0.99821 0.17032 0.15449
Table 4.1b: Goodness of Fit

<table>
<thead>
<tr>
<th>Number of Prior Purchases</th>
<th>Data</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.00286</td>
<td>0.00280</td>
</tr>
<tr>
<td></td>
<td>(0.00016)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.01747</td>
<td>0.01525</td>
</tr>
<tr>
<td></td>
<td>(0.00144)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.03501</td>
<td>0.04179</td>
</tr>
<tr>
<td></td>
<td>(0.00375)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.04530</td>
<td>0.06402</td>
</tr>
<tr>
<td></td>
<td>(0.00591)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.08566</td>
<td>0.09323</td>
</tr>
<tr>
<td></td>
<td>(0.01207)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.07368</td>
<td>0.11427</td>
</tr>
<tr>
<td></td>
<td>(0.01198)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.12385</td>
<td>0.13107</td>
</tr>
<tr>
<td></td>
<td>(0.02231)</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.16547</td>
<td>0.15450</td>
</tr>
<tr>
<td></td>
<td>(0.03151)</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.13291</td>
<td>0.17667</td>
</tr>
<tr>
<td></td>
<td>(0.02700)</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.28378</td>
<td>0.18147</td>
</tr>
<tr>
<td></td>
<td>(0.05240)</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.20879</td>
<td>0.20961</td>
</tr>
<tr>
<td></td>
<td>(0.04260)</td>
<td></td>
</tr>
</tbody>
</table>

Note: Standard deviations of data probabilities in parentheses
Table 4.2: Additional Estimates

| Parameter | Myopic w/ Flexible Prestige Advertising | Myopic w/ Random Coefficient on Prestige Advertising | Myopic w/ multi-period learning (|\theta|_{11} = 1) |
|-----------|----------------------------------------|-------------------------------------------------|-------------------------------------------------|
| \theta_1 - Price | -5.5258 (0.33081) | -5.6228 (0.32840) | -5.7248 (0.33384) |
| \theta_2 - Store Coupon | 3.12930 (0.81223) | 3.06820 (0.82226) | 3.11890 (0.81513) |
| \theta_3 - Prestige Advertising | -0.00803 (0.05651) | -0.03285 (0.04685) | -0.07276 (0.02448) |
| \theta_4 - Competitor’s Price | -0.69068 (0.22560) | -0.71465 (0.22015) | -0.7276 (0.22448) |
| \theta_5 - Time Trend on Outside Alternative | 1.23630 (0.16740) | 1.07650 (0.17547) | 1.28130 (0.19354) |
| \sigma_i - Variance of the distorted \theta | 1.7547 (0.16740) | 1.48520 (0.17547) | 2.30260 (0.19354) |
| around the \theta | 1.45320 (0.14222) | 1.70980 (0.14617) | 4.90710 (2.59952) |
| \sigma_j - Consumer’s perceived variance of the distorted \theta | 1.45320 (0.14222) | 1.70980 (0.14617) | 4.90710 (2.59952) |
| \rho - Correlation Coefficient of the inflated Information Cost of Yoplait 150 | 0.62982 (0.04930) | 0.74675 (0.02701) | 0.79062 (0.03428) |
| \delta - Mean Experience Quality of Yoplait | -0.75524 (0.09484) | -0.84118 (0.07672) | 1.72740 (0.15174) |
| Discount Factor | 0 | 0 | 0 |
| \lambda_i - Constant | -3.34571 (0.44905) | -3.2320 (0.85789) | -3.01320 (1.30272) |
| \sigma \lambda | 1.76680 (0.15519) | 1.71530 (0.11035) | 1.73290 (0.12572) |
| \lambda_i^2 m_i^\alpha | 0.15037 (0.09829) | 0.06683 (0.14448) | 0.19758 (0.15790) |
| \lambda_i^2 m_i^\alpha < 2 | 0.06683 (0.14448) | 0.19758 (0.15790) | 0.23300 (0.16749) |
| \lambda_i^2 m_i^\alpha < 3 | 0.23300 (0.16749) | 0.04442 (0.21307) | -0.04442 (0.21307) |
| \lambda_i^2 m_i^\alpha < 6 | -0.04442 (0.21307) | 0.04442 (0.21307) | 0.04442 (0.21307) |
| \lambda_i^2 m_i^\alpha < 5 | 0.23300 (0.16749) | 0.04442 (0.21307) | 0.04442 (0.21307) |
| S.D. of Random Coefficient on \theta | 0.11258 (0.07562) | 0.11258 (0.07562) | 0.11258 (0.07562) |
| Log Likelihood | -3940.0390 | -3939.6724 | -3921.7334 |

Note: Below parameters not estimated, standard errors obtained by delta method.

| Informative Advertising | 0.34616 (0.03477) | 0.38097 (0.17062) | 0.33471 (0.23470) |
| Effect - 1 Week | 0.38876 (0.04632) | 0.41881 (0.18613) | 0.38903 (0.21620) |
| Informative Advertising | 0.41421 (0.05497) | 0.44082 (0.20523) | 0.62047 (0.23471) |
| Effect - 20 Weeks | 0.41421 (0.05497) | 0.44082 (0.20523) | 0.62047 (0.23471) |

Note: Standard Errors in parentheses. In columns 2 and 3 these are not adjusted for simulation error. Because of different simulation methods, likelihoods in columns 2 and 3 not directly comparable to those in column 1 and Table 4.1. Not all \lambda_i terms shown.
Table 5.1: Yoplait 150 Welfare Results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimated Equilibrium</th>
<th>Ad Ban Adjusting Price</th>
<th>Ad Ban w/o Adjusting Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Price</td>
<td>0.6527</td>
<td>0.6438</td>
<td>0.6527</td>
</tr>
<tr>
<td>Marginal Cost</td>
<td>0.4221</td>
<td>0.4221</td>
<td>0.4221</td>
</tr>
<tr>
<td>Total Sales</td>
<td>236371.00</td>
<td>225498.00</td>
<td>218377.00</td>
</tr>
<tr>
<td>Total Revenue</td>
<td>148360.93</td>
<td>140129.22</td>
<td>137047.13</td>
</tr>
<tr>
<td>Production Costs</td>
<td>99762.744</td>
<td>95173.686</td>
<td>92168.197</td>
</tr>
<tr>
<td>Advertising Costs</td>
<td>9329.996</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total Costs</td>
<td>100092.74</td>
<td>95173.686</td>
<td>92168.197</td>
</tr>
<tr>
<td>Profits</td>
<td>39268.186</td>
<td>44955.529</td>
<td>44878.933</td>
</tr>
<tr>
<td>Compensating Variation</td>
<td>45541.838</td>
<td>46554.529</td>
<td>44962.923</td>
</tr>
<tr>
<td>Total Welfare</td>
<td>84810.024</td>
<td>91510.058</td>
<td>89841.856</td>
</tr>
</tbody>
</table>
Figure 5.1

Prices and Marginal Costs

Experience Quality

Marginal Cost
Mean Price
Mean Price w/Ad Ban
Figure 5.2

Compensating Variation

Experience Quality

CV
CV w/Ad Ban
CV w/Ad Ban (No Price Change)
Figure 5.2b

Decomposition of Effects of Ban on Consumers

- Value of Advertising Information
- Value of Equilibrium Price Adjustment (Ad to No Ad)
- Total Change in CV (Ad to No Ad)
Figures 5.3 and 5.4

Sales

Revenue

Experience Quality

Sales

Sales w/ Ad Ban

Revenue

Revenue w/ Ad Ban
Figures 5.5 and 5.6

**Costs**
- Production Costs
- Advertising Costs
- Total Costs
- Production (= Total) Costs w/Ad Ban

**Profits**
- Profits
- Profits w/o Ad Costs
- Profits w/Ad Ban
References


*Advertising Age*, Serial, Chicago, Ill., Crane Communications


Schmalensee R. (1972) *The Economics of Advertising* Amsterdam, North Holland Publishers


