Advertising, Learning, and Consumer Choice in Experience Good Markets: An Empirical Examination
Author(s): Daniel A. Ackerberg
Reviewed work(s):
Source: *International Economic Review*, Vol. 44, No. 3 (Aug., 2003), pp. 1007-1040
Published by: Blackwell Publishing for the Economics Department of the University of Pennsylvania and Institute of Social and Economic Research -- Osaka University
Stable URL: http://www.jstor.org/stable/3663546
Accessed: 23/11/2011 13:49

INTERNATIONAL ECONOMIC REVIEW
Vol. 44, No. 3, August 2003

ADVERTISING, LEARNING, AND CONSUMER CHOICE IN EXPERIENCE GOOD MARKETS: AN EMPIRICAL EXAMINATION*

BY DANIEL A. ACKERBERG[1]

Economics Department, University of California, Los Angeles and NBER

This article empirically analyzes different effects of advertising in a nondurable, experience good market. A dynamic learning model of consumer behavior is presented in which I 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.

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 nondurable 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.

* Manuscript received July 1998; revised April 2002.

[1] This article 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, the Editor Ken Wolpin, two anonymous referees, 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. Please address correspondence to: Daniel A. Ackerberg, Economics Department, University of California, Los Angeles, Box 951477, Los Angeles, CA 90095-1477, USA. Phone: (310) 794-5427. Fax: (310) 825-9528. E-mail: *ackerber@econ.ucla.edu.*

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, I 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. I 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.[2] 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 noninformative 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 (1977) 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. My 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

[2] 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, p. 289). On the other hand, prestige effects might increase barriers to entry by creating product differentiation and market power.

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 observable and unobservable characteristics, he argues that they should not be affected by exposures to informative advertising.[3] In contrast, he hypothesizes that Becker-like prestige or image effects of advertising should generally affect both inexperienced and experienced users of the brand.[4] 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, I formally model consumer information, introducing a model of consumer behavior that explicitly includes both informative and prestige effects of advertising. My model is similar to Eckstein et al.'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 nondurable, 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. My Becker-like prestige or image effect of advertising enters directly in the utility function, influencing utility independently of beliefs on inherent product characteristics.[5] My resulting model of advertising is closest to the model estimated by Erdem and Keane (1996). They also develop a dynamic model of informative advertising, but examine the demand implications of this single effect and do not distinguish different effects of advertising.[6]

[3] 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 nondurables like those considered here. See Ackerberg (2001) for other exceptions and more discussion.

[4] 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*.

[5] I stress that these prestige/image effects constitute completely rational behavior on the part of consumers. Terming 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.

[6] There are other significant differences between the two models. One is the extent of consumer heterogeneity. My 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 way that informative advertising is modelled and in the policy analysis that is performed. They examine alternative firm advertising strategies while I measure the value of information in advertising.

The simplest representation of the important empirical components of my 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., iff

$$E[U(\delta, a) \mid a] > k$$

The utility function U contains δ, 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 δ 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 my prestige or image effect of advertising—advertising influencing utility *given* inherent product characteristics. Secondly, the expectation over δ is conditioned on a. This is my informative effect of advertising—I allow advertising to "tell" the consumer something about the brand's characteristics δ. 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, my 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 my data set. I therefore choose just one informative effect to include in my structural model, that of signaling. Reasons for this choice include: (1) the 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 my market. However, I believe that these different informative effects of advertising should in some sense be observationally equivalent in my data: all tend to affect inexperienced rather than experienced consumers. As a result, I feel that my conclusions regarding significance or insignificance of my informative and prestige effects would not substantially change if I had instead modeled one of the other informative effects. In summary, I 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., "prior experience," shift means and not variances. This contrasts with my 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.[7] Not only will ignorance of this dispersion be inefficient, but it can potentially generate spurious results.[8] 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. I stress that, unlike my main empirical conclusions, the welfare analysis I perform is probably highly dependent on my choice to model informative advertising as a signaling effect. Though this limits the applicability of the welfare results, I feel that it is still an interesting and enlightening exercise.[9]

1.3. *Summary of Results.* Estimates of my 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 insignificant prestige effect of advertising. This supports the reduced form conclusion that the advertisements

[7] 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 article.

[8] For example, consider a situation where both experienced and inexperienced consumers have $E[U(\delta, a) \mid a]$s (EUs) distributed around zero (assume consumers purchase if $EU > 0$), but experienced consumers' EUs have more dispersion (a higher variance). In this case, a burst of prestige advertising that shifts *all* consumers' EUs 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.

[9] See papers by Miller (1984), Wolpin (1984), Pakes (1986), and Eckstein and Wolpin (1991) for further discussion on the advantages of structural models.

in this data primarily affected consumers through the provision of information. Under the strong assumption that this is in fact signaling information, my 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 my general model of consumer behavior and Section 3 describes the data used in this study. Section 4 details my empirical specification and presents my results. In Section 5, I perform my welfare experiments and Section 6 concludes.

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 nondurable 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 nondurable 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:

$$\max_{c_\tau(I_{i\tau})\tau>=t} \mathrm{E}\left[\sum_{\tau=t}^{\infty} \beta^{\tau-t} U_{ic_\tau\tau} \mid I_{it}\right] \tag{2.1}$$

where $c_t \in \{1,2\}$ is the consumer's choice at t (2 represents the outside alternative) and β is the per-period discount factor.[10]

As is now relatively common in the empirical analysis of differentiated products, I 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, I assume the utility consumer i obtains from purchase and consumption of brand j in period t is

$$U_{i1t} = \lambda_i + \theta_1 p_{it} + \delta_{it+1} + \theta_3 m_{it}^a + \epsilon_{it} \tag{2.2}$$

λ_i represents consumer i's tastes for the *observable* characteristics of the new product. This will generally be a function of the products observable characteristics and the consumer's known tastes for these characteristics. ϵ_{it} represents idiosyncratic,

[10] 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.

time-varying shocks to the utility the consumer derives from consuming that are *known* prior to the purchase decision. Though I defer its formal definition until later, m_{it}^a is a measure of what consumer i currently knows about how much the brand is advertising. Its entry into the utility function will represent my 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 δ_{it+1}, which I 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, δ_{it+1} might capture how the brand actually tastes to the consumer (conditional on its observed characteristics that enter λ_i). Although δ_{it+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 δ_{it+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 δ_{it+1} and knows its value for all future t.

As in Eckstein et al. (1988) and Erdem and Keane (1996), I 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

$$\delta_{it+1} = \delta_i + \nu_{it+1} \quad \text{where} \quad \nu_{it+1} \sim \text{i.i.d. } N(0, \sigma_\nu^2) \tag{2.3}$$

Although δ_{it+1} is realized (observed) by the consumer after consumption, its components, δ_i and ν_{it+1}, are never individually observed. δ_i is the mean experience utility consumer i obtains from the brand. ν_{it+1} are i.i.d. confounding variables that cannot be distinguished from this mean. In the case of food products, variance in ν_{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.[11] In contrast to the i.i.d. ν_{it+1}, δ_i is persistent over time. It is thus beneficial for the consumer to use information contained in observed δ_{it+1}s to learn about its value. In the degenerate case where $\sigma_\nu^2 = 0$, we have the one-consumption learning process described above where δ_i (and thus δ_{it+1} $\forall t$) is learned after one consumption experience. In the nondegenerate case, consumption and subsequent realization of δ_{it+1} does not exactly reveal δ_i, but it does provide information about it. This information acquisition will be consistently modeled in a Bayesian learning framework.

[11] In all these cases the important thing is that the ν_{it} are indistinguishable from δ_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.

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

$$a_{it} = a + \xi_{it} \quad \text{where} \quad \xi_{it} \sim \text{i.i.d. } N(0, \sigma_\xi^2) \tag{2.4}$$

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.[12] Although consumers do not directly observe a brand's mean advertising intensity a, I 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 δ_i, 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.[13]

I consistently model information provided by the observed a_{it}s and δ_{it+1}s on the relevant unknowns a and δ_i as a bivariate Bayesian learning process. In matrix notation, Equations (2.3) and (2.4) become

$$\begin{pmatrix} \delta_{it+1} \\ a_{it} \end{pmatrix} \sim \text{i.i.d. } N\left(\begin{pmatrix} \delta_i \\ a \end{pmatrix} \Phi \right) \quad \text{where} \quad \Phi = \begin{bmatrix} \sigma_\nu^2 & 0 \\ 0 & \sigma_\xi^2 \end{bmatrix} \tag{2.5}$$

The assumed diagonality of Φ implies that there is no correlation between δ_{it+1} and a_{it} *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 δ_i:

$$\text{Initial prior:} \quad \begin{pmatrix} \delta_i \\ a \end{pmatrix} \sim N\left(m_0 = \begin{pmatrix} 0 \\ m_0^a \end{pmatrix}, \Sigma_0 \right) \tag{2.6}$$

[12] In my empirical work, I generalize to having advertising exposures distributed around an individual specific mean (i.e., $a_{it} = a_i + \xi_{it}$).

[13] As discussed in the introduction, complexity and identification issues necessitated the inclusion of only one informative effect of advertising in my model. There are clearly alternative specifications. Erdem and Keane's (1996) similar dynamic model has advertising explicitly informing the consumer on δ_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)).

generates a learning process in which a consumer's posterior on brand j after a history of observed advertising intensities, $\{a_{i1}, \dots, a_{it}\}$, and consumption experiences, $\{\delta_{i1}, \dots, \delta_{iK_{it}}\}$, is given by[14]

$$\begin{pmatrix} \delta_i \\ a \end{pmatrix} \sim N(m_{it}, \Sigma_{it}) \tag{2.7}$$

where

$$m_{it} = \begin{pmatrix} m_{it}^{\delta} \\ m_{it}^{a} \end{pmatrix} = (\Sigma_0^{-1} + \eta_{it}\Phi^{-1})^{-1}(\Sigma_0^{-1}m_0 + \eta_{it}\Phi^{-1}\overline{z}_{it})$$

$$\Sigma_{it} = (\Sigma_0^{-1} + \eta_{it}\Phi^{-1})^{-1} \quad \overline{z}_{it} = \begin{pmatrix} \frac{1}{K_{it}}\sum_{k=1}^{K_{it}} \delta_{ik} \\ \frac{1}{t}\sum_{\tau=1}^{t} a_{i\tau} \end{pmatrix} \quad \text{and} \quad \eta_{it} = \begin{bmatrix} K_{it} & 0 \\ 0 & t \end{bmatrix}$$

and where 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 λ_i and δ_{it+1}, setting the initial prior mean on δ_i equal to 0 is a normalization. Essentially, I am treating the expected value of the unobserved characteristic as an observable characteristic (i.e., it is part of λ_i).

m_{it} is a weighted average of initial priors and observed realizations of δ_{it+1} and a_{it}. An important result of Bayesian learning is that these posterior means and variances summarize all the consumer's information on δ_i and a. Thus, the current posterior (m_{it}, Σ_{it}) is sufficient to define perceived distributions over both future δ_{it+1}'s and a_{it}'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 Σ_0 is zero, then the learning processes on δ_i and a are independent. On the other hand, a nonzero covariance term indicates a perception by consumers that δ_i 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 δ_i.

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

$$a = \beta_0 + \beta_1\delta$$

where δ is the brand's mean experience utility level over the population. Then, assuming (1) a normal population distribution of δ_i around $\delta(\delta_i \sim N(\delta, \sigma_i^2))$ and (2) a normal prior on $\delta(\delta \sim N(0, \sigma_j^2))$ a Bayesian consumer's initial prior variance

[14] The derivation of the following conjugate result is a fairly simple extension of the derivation for a multivariate normal with *equal* draws in DeGroot.

matrix is

$$\text{Initial prior:} \quad \begin{pmatrix} \delta_i \\ a \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ \beta_0 \end{pmatrix}, \begin{bmatrix} \sigma_i^2 + \sigma_j^2 & \beta_1 \sigma_j^2 \\ \beta_1 \sigma_j^2 & \beta_1^2 \sigma_j^2 \end{bmatrix}\right)$$

Thus, in a signalling equilibrium where $\beta_1 > 0$, consumers interpret high levels of advertising as a signal of a higher δ_i.[15]

In addition to this informative effect of advertising, I accommodate prestige or image effects of advertising by allowing the consumer's current posterior mean on a, m_{it}^a, to enter directly into the utility function (2.2). As prestige and image effects of advertising are somewhat nontangible, this term warrants some discussion. One alternative is to interpret this term structurally, i.e., all else equal (in particular, expectations over δ_{it+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, "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 "advertising-based prestige."[16]

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 do not know that American basketball player 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, my coefficient on m_{it}^a 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}^a entering directly into the utility function.[17] Clearly I am in somewhat murky waters regarding the specification of these image and prestige effects. To partially

[15] I take the view that consumers are rational in their beliefs, i.e., a positive covariance term in priors is actually generated by a signaling equilibrium. An alternative interpretation is that I am simply estimating consumer beliefs, without assuming anything about where they come from or whether or not they are correct.

[16] Note that when I discuss advertising's effect on prestige, I am 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. My hope is that these other aspects of prestige are constant across time and known to consumers, and thus captured by my constant term and consumer random effects.

[17] I am assuming that each television advertisement portrays the same images, i.e. that advertising copy is the same or similar across commercials. Given that I do not observe advertising copy, this is likely all I can do.

compensate, in empirical work I try to be as general as possible with the specification. However, I also have to admit the possibility that this representation might pick up misspecifications

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}, Σ_{it}) is sufficient to define all the consumer's current information on δ_i and a, the sequential maximization problem of (2.1) can be transformed into the following Bellman's equation:

$$\begin{aligned}(2.8)\quad V_i(p_{it}, m_{it}, \Sigma_{it}, \epsilon_{it}) = \max_{c_{it}\in\{1,2\}} & E\big[U\big(p_{it}, \lambda_i, \epsilon_{it}, m^a_{it}, \delta_{it+1}\big) \\ & + \beta V_i\big(p_{it+1}, m_{it+1}, \Sigma_{it+1}, \\ & \epsilon_{it+1}\big) \mid (p_{it}, m_{it}, \Sigma_{it}, \epsilon_{it}), c_{it}\big]\end{aligned}$$

where the state space $(p_{it}, m_{it}, \Sigma_{it}, \epsilon_{it})$ contains prices, the current posterior, and time-varying preference shocks ϵ_{it}. The expectation is over current period experience utility δ_{it+1} as well as next period's state. For consumer i with posterior (m_{it}, Σ_{it}) facing prices p_{it} and shocks ϵ_{it}, $V_i(\cdot)$ is the perceived expected discounted value of future utilities. This value function has an associated policy function, $c_{it} = c_i(p_{it}, m_{it}, \Sigma_{it}, \epsilon_{it})$, 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, I have used numeric solution methods to solve and generate predictions of the model. I 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 ($\sigma_\nu^2 = 0$, $[\Sigma_0]_{11} > 0$), purchase patterns change after the first purchase but not thereafter. If there is variance in δ_{it+1} ($0 < \sigma_\nu^2 < \infty$), purchasing patterns do change after future purchases, the extent and length depending on σ_ν^2 and $[\Sigma_0]_{11}$. On the other hand, if there is no learning ($\sigma_\nu^2 = \infty$ or $[\Sigma_0]_{11} = 0$), 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 δ_i. However, while consumption provides direct information on δ_i, 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 I take to the data to distinguish between the two effects of advertising.

3. THE DATA

I 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.[18] In each of two geographically isolated markets (Sioux Falls, South Dakota and Springfield, Missouri), shopping trips and purchases of approximately 2000 households at 80% of area supermarkets and drugstores were followed for three years (1986–88). There is also data on weekly prices at each store, so we essentially know prices on each household's shopping trips.[19] 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.

[18] This type of data has primarily been analyzed in the marketing literature (e.g., Guadagni and Little, 1983; Pedrick and Zufryden, 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.

[19] When a consumer purchases a product, we observe the exact transaction price. To measure prices when a product was not 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, my focus is on the effects of advertising, not price elasticities.

TABLE 1

SUMMARY STATISTICS

Variable	Market 1	Market 2
Households	950	825
Total shopping trips	67,051	54,308
Average shopping trips per household	70.58	65.82
	(33.39)	(31.82)
Average price of Yoplait 150 (cents)	0.645	0.663
	(0.060)	(0.079)
Shopping trips with Yoplait 150 purchase	302	656
Manufacturers coupons redeemed for Yoplait 150	16	238
Shopping trips with other yogurt purchase	5432	3863
Households trying Yoplait 150	123	184
Households trying other yogurts	648	512
Commercial exposures	12,918	12,563
Commercial exposures per household	13.60	15.22
	(10.81)	(9.96)
Advertising share of Yoplait 150	0.35	0.37
Market share of Yoplait 150	0.05	0.14

NOTE: Standard errors in parentheses where applicable.

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, I again focus specifically on Yoplait 150, modeling competing brands in a informationally static and sparse framework.

Table 1 gives some summary statistics for the data following Yoplait 150's introduction.[20] 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. I urge the reader to consult Ackerberg (1997, 2001)

[20] Only households whose television viewing was recorded are included both here and in estimation. I 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.

for 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[21] and advertising.[22]

4. ESTIMATION

We now move to estimation of my model using this data, starting with a more detailed discussion of my empirical specification, estimation issues, and numeric techniques used in solving the dynamic programming problem. Because the empirical model is fairly complicated and nonlinear, I intuitively discuss how my 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, I assume that the time frame of my model is the consumer (i.e., household) shopping trip. Specifically, I 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.[23] We specify our consumers' single period utility

[21] 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, I 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 I do include this as an explanatory variable.

[22] 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. I 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 my 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).

[23] In preliminary specifications I compared these two-choice models to three-choice models (with the choices: Yoplait 150, a different brand of yogurt, or no yogurt) and obtained similar results. Note that I also completely ignore the number of yogurts purchased on a particular shopping occasion, avoiding what is in actuality a more complicated discrete/continuous choice. Regarding the learning process, my assumption is that purchase quantities and variables possibly affecting the length of the learning process (in particular, family size) scale together so that one purchase *occasion* provides the same information across households.

functions as

$$U_{it} = \begin{cases} U_{i1t} = \lambda_i + \theta_1 p_{it} + \theta_2 sc_{it} + \delta_{it+1} + \theta_3 m_{it}^a + \epsilon_{i1t} & \text{if Yoplait 150 purchased and consumed at } t \\ U_{i2t} = \theta_4 p_{it}^{\text{oth}} + \epsilon_{i2t} & \text{otherwise (outside alternative)} \end{cases}$$

The variables p_{it}, sc_{it}, and p_{it}^{oth} measure, the price of Yoplait 150, the value of a possible store coupon available for Yoplait 150, and a scalar measure of other yogurts' prices,[24] 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 θ_1, θ_2, and θ_4 measure marginal effects of these variables on utility.

λ_i is modeled as a linear combination of observable consumer characteristics (y_i) plus a normally distributed random variable with variance σ_λ^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.[25] The "random effect" component of λ_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, ϵ_{i1t} and ϵ_{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 my additive normalization; the fixed variance of the ϵ_{it}s the multiplicative.

In my model, consumers want to learn how much advertising a brand is engaging in. Thus, I 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 of television watched during that period.[26] 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 $t-1$ and t. Though this directly controls for

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

[25] This "presample" purchase data is assumed exogenous to my model, and as might be expected are very good predictors of λ_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. λ_i represents the sum of the utilities from these characteristics for consumer i.

[26] There are a number of reasons for this definition. First, I believe it best corresponds to the specific effects of advertising included in the model. Second, it may partially alleviate measurement error resulting from TV meter problems. Third, Ackerberg (2001) found intensities to do better at explaining the data than absolute number of advertisements.

consumers differing in *how much* television they watch, the data also indirectly suggest that consumers persistently differ in *what* they watch (I find statistically significant differences in consumers' mean (over time) a_{it}s). To accommodate this, I add an additional level of variance to the advertising exposure process. Specifically, I assume that the a_{it}s are distributed normally around a consumer-specific advertising intensity a_i (measuring *what* consumer i watches) which in turn are distributed normally around the brand's advertising intensity a.[27] The variance of a_i around a, σ_a^2, is taken directly from the data. In addition, a is estimated directly from the data. This implies that we need not estimate the parameter β_0, since conditional on the parameters β_1 and δ, my signalling equilibrium equation implies $\beta_0 = \beta_1\delta - \widehat{a}$.

4.2. *Identification.* It is important to discuss how these learning parameters are identified 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 patterns will converge to some "postinformation" level. σ_λ^2, the variance in the unobserved component of consumers' *known* tastes for Yoplait 150, is identified by unobserved heterogeneity in consumers' "preinformation" (pre-first-purchase) behavior. On the other hand, σ_i^2, the variance of the unknown taste δ_i across the population, is identified by comparing the variance of "preinformation" heterogeneity to the variance of "postinformation" heterogeneity. δ, the mean experience utility of Yoplait 150, is assessed by a comparison of the means of these two distributions, i.e., whether "postinformation," consumers (on average) purchase Yoplait 150 more or less than "preinformation" (net of experimentation behavior due to dynamic optimization). σ_v^2, the per-period variance in experience utility, is identified by the number of consumption experiences it takes for consumers to learn δ_i, i.e., how many consumption experiences it takes for purchasing patterns to converge to the "postinformation" level. If, for example, purchase patterns change after initial purchases, but not thereafter, it is indicative that $\sigma_v^2 = 0$, i.e., a one-consumption learning process. The advertising-related coefficients, β_1 and θ_3, are identified by the effects of advertising exposures on inexperienced and experienced consumers (both in an absolute sense (θ_3) and relatively (β_1)).[28]

[27] As the a_{it} are observables this does nothing to estimation or the model except that now m_{it}^a is the consumer's posterior on a_i (rather than on a). Literally, this changes slightly the interpretation of the "prestige" effect as now it's a high belief on a_i that generates utility. Practically, as consumers are using the same information (a_{it}s) to learn about a and a_i, the posteriors on a and a_i are highly correlated.

[28] The last "learning process" parameter, σ_j^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).

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 δ_i, his random component of λ_i, and his realizations of experience utility (δ_{it+1}) at each purchase occasion. Recall that these unobservables are assumed mutually uncorrelated except for the fact that experience utility realizations are distributed around δ_i. In addition, these unobservables are assumed independent of my observables $y_i, a_{it}, p_{it}, sc_{it,}$ and p_{it}^{oth}.[29] Because of the persistent unobservables and the dependence of purchase probabilities on lagged endogenous variables (through posteriors), I 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.[30] This results in the following likelihood function (for consumer i):

$$
\begin{aligned}
L_i(\theta) &= \Pr\left[\left\{c_{it} = c\left(m_{it}\left(\delta_{it}^t, a_{it}^t, c_{it}^{t-1};\theta\right), z_{it}, \Sigma_{it}, \lambda_i, \epsilon_{it};\theta\right)\right\}_{t=1}^{T_i} \mid z_{it}^{T_i}, a_{it}^{T_i}, y_i\right] \\
&= \int \Pr\left[\left\{c_{it} = c\left(m_{it}\left(\delta_{it}^t, a_{it}^t, c_{it}^{t-1};\theta\right), z_{it}, \Sigma_{it}, \lambda_i, \epsilon_{it};\theta\right)\right\}_{t=1}^{T_i} \mid z_{it}^{T_i}, a_{it}^{T_i}, \lambda_i, \delta_{it}^{T_i}\right] \\
&\qquad p\left(d\delta_{it}^{T_i} \mid \delta_i;\theta\right) p(d\delta_i \mid \theta) p(d\lambda_i \mid y_i;\theta) \\
&= \int \prod_{t=1}^{T_i} \Pr\left[c_{it} = c\left(m_{it}\left(\delta_{it}^t, a_{it}^t, c_{it}^{t-1};\theta\right), z_{it}, \Sigma_{it}, \lambda_i, \epsilon_{it};\theta\right) \mid z_{it}, a_{it}^t, \lambda_i, \delta_{it}^t, c_{it}^{t-1}\right] \\
&\qquad p\left(d\delta_{it}^{T_i} \mid \delta_i;\theta\right) p(d\delta_i \mid \theta) p(d\lambda_i \mid y_i;\theta)
\end{aligned}
$$

where c_{it} is the consumer's observed choice in period t, $c(\cdot)$ is the model's predicted choice, $z_{it} = (p_{it}, p_{it}^{\text{oth}}, sc_{it})$, T_i is the total number of shopping trips of consumer i, and superscripts indicate histories of a variable through that point (e.g., $a_{it}^t = \{a_{i1}, \ldots, a_{it}\}$). The $\Pr[\cdot]$ in the last line is the probability that the period t logit errors (ϵ_{it}) are such that the model's predicted choice equals our observed choice, conditional on λ_i, past choices c_{it}, and past realized δ_{it+1}s. The predicted choice function $c(\cdot)$ 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 ϵ_{it}s, the last $\Pr[\cdot]$ has a closed form solution in both the myopic and fully dynamic cases (Rust, 1987).

[29] 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 ϵ_{i1t} 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 toward consumers who like Yoplait 150 more than our observables predict (i.e., a_{it} is correlated with the random component of λ_i or δ_i). Ackerberg (2001) did not find statistical support for this possibility.

[30] With the exception of the logit errors, the integrals generated by these unobservables are not analytically computable and I 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.

4.4. *Dynamic Programming Solution.* In estimating the full dynamic model, we must solve the consumer's dynamic programming problem to obtain $c(\cdot)$. 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. My utility specification and the assumed learning process generate the Bellman's equation:

$$V(s_{it};\theta) = \max\{E[U_{i1t} + \beta V(s_{it+1};\theta) \mid s_{it}, c_{it} = 1]$$
$$U_{i2t} + \beta E[V(s_{it+1};\theta) \mid s_{it}, c_{it} = 2]\}$$

where the state space $s_{it} = (\lambda_i, m_{it}, \Sigma_{it}, p_{it}, sc_{it}, p_{it}^{\text{oth}}, \epsilon_{iot}, \epsilon_{i1t})$. Although this state space appears to be quite large, there are a number of simplifications and assumptions that I use to significantly reduce the dimensionality of the problem and allow for relatively quick numeric solution. First, conditional on Σ_0 and Φ, 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 Σ_{it}. Therefore, my assumption that Σ_0 and Φ are constant across consumers[31] allows us to replace Σ_{it} in the state space by K_{it} and TV_{it}. Second, because δ_{it+1} enters the utility function linearly, λ_i can be merged into the learning process (so my consumer Bayesian updates on the sum $\delta_i + \lambda_i$). 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. ϵs from the effective state space, and as we assume that consumers perceive p_{it}, sc_{it},and p_{it}^{oth} to be nonserially correlated,[32] we end up with a four-dimensional problem where $\overline{s_t}$, 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 ϵs 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_{it}^{oth}, implies that to compute the expectations in the value function we need only numerically integrate over the distribution of next period's posterior means.

Because the elements of $\overline{s_t}$ are either continuous variables (m_{it}) or take on a large number of discrete values ($K_{it}$and TV_{it}), the state space must be discretized

[31] For Φ to be constant, I 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, Φ 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 these would at the very least add an extra state variable to the model.

[32] I assume a four-point (estimated from the data) perceived distribution of future p_{it}, and a degenerate distribution of sc_{it} and p_{it}^{oth} (i.e., consumers expect that next shopping trip, sc_{it} and p_{it}^{oth} will be at their respective means). The assumption that the p_{it} are i.i.d. 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 does not enter the myopic estimation results at all).

in order to apply the method of successive approximations and numerically solve the above dynamic programming problem. I do not discretize the entire dynamic problem, but choose points at which I will solve for the (approximate) value function. In the following estimation results I have discretized the state space into 10 to 20 points in each dimension. Because the numerical integration mentioned above is only two dimensional I chose to use quadrature rather than Monte Carlo. Since the quadrature points generate future states that are not my discretized ones, I use linear interpolation to evaluate the value function at these states.

4.5. *Results.* Table 2 presents maximum likelihood estimates of the above model. In initial runs I had trouble obtaining reasonable estimates of σ_v^2, the per-period variance in experience utility. My 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 my estimates assume a one-period learning process ($\sigma_v^2 = 0$). This assumption has the added benefit of greatly reducing both the computational burden of the dynamic programming problem and likelihood evaluation. I have capitalized on this computational reduction to increase the precision and accuracy of my discretization and integral evaluation over what would have otherwise been possible.[33]

The first two columns of Table 2 contain results under the myopic assumption on behavior, with and without a time trend[34] on the outside alternative. The last three columns are results from estimation of the full dynamic model. In all five 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 σ_λ indicates that there is significant unexplained heterogeneity in consumers' *initial* valuations of Yoplait 150. The very large significance of the estimates of σ_i strongly support the existence of imperfect information and learning, indicating that consumers' have

[33] 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}^{δ}- 20 points, m_{it}^{a}- 10 points, K_{it}- 2 points (either purchased or never purchased), TV_{it}- 10 points), 49 draws on the Bellman's equation integral (for a total of 196,000 three-dimensional interpolations per contraction iteration), and each of the distributions of δ_i and the random component of λ_i discretized into a 26-point normal for likelihood evaluation (for a total of 262 draws), one function evaluation with a discount rate of 0.98 on an UltraSparc 167 MHz processor in highly optimized C code takes about 7 minutes (~5.5 minutes for the contraction mapping to converge, 1.5 minutes for likelihood evaluation).

[34] 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.

TABLE 2

MYOPIC AND FULL DYNAMIC ESTIMATES

Parameter	Myopic Model	Myopic Model w/ Time Trend	Dynamic Model	Dynamic Model w/ Time Trend	No Prestige Advertising
θ_1 (Price)	−5.26140	−5.54170	−5.29690	−5.49230	−5.48930
	(0.31620)	(0.32557)	(0.31454)	(0.33230)	(0.32980)
θ_2 (Store coupon)	3.11930	3.11540	3.04760	3.11030	3.09590
	(0.82961)	(0.80984)	(0.83679)	(0.81199)	(0.80927)
θ_3 (Prestige advertising)	−0.10537	0.00281	−0.15855	−0.02469	0
	(0.03751)	(0.03886)	(0.04117)	(0.04415)	
θ_4 (Competitor's price)	−0.74010	−0.69667	−0.77747	−0.69536	−0.70704
	(0.22169)	(0.22243)	(0.22154)	(0.22136)	(0.21827)
θ_5 (Time trend on outside alternative)	0	1.16370	0	0.94299	0.98856
		(0.17081)		(0.16134)	(0.13694)
σ_i (Variance of δ_i around δ)	1.76690	1.77900	1.86030	1.83750	1.81620
	(0.13582)	(0.13985)	(0.13924)	(0.13319)	(0.13261)
σ_j (Consumer's percieved variance of δ)	0.92348	1.65730	1.88830	0.64559	0.59278
	(0.23840)	(1.57685)	(1.33743)	(0.32074)	(0.27870)
ρ (Correlation coefficient of Σ_0, informative advertising)	0.36563	0.67287	0.14273	0.13258	0.12347
	(0.12444)	(0.36086)	(0.09219)	(0.04054)	(0.03619)
δ (Mean experience quality of Yoplait 150)	−1.24190	−0.71716	0.67938	0.89500	0.89878
	(0.72998)	(0.99348)	(0.78308)	(0.28409)	(0.27600)
Discount factor	0	0	0.98139	0.98	0.98
			(0.01885)		
λ_i (Constant)	−4.45980	−3.41110	−4.88810	−4.51520	−4.60370
	(0.97585)	(1.08139)	(0.53794)	(0.46500)	(0.46387)
λ_i (Market dummy)	1.65010	1.49190	1.22720	1.25530	1.27340
	(0.19015)	(0.17476)	(0.23433)	(0.17639)	(0.17298)
λ_i (Income)	0.08351	0.07467	0.05884	0.06475	0.05995
	(0.03342)	(0.03114)	(0.02958)	(0.02679)	(0.02658)
λ_i (Family size)	−0.07470	−0.06929	−0.08792	−0.06254	−0.02484
	(0.08044)	(0.07055)	(0.06345)	(0.06048)	(0.06061)
λ_i (Presample yogurt purchases)	0.01494	0.01380	0.01185	0.01148	0.01087
	(0.01485)	(0.01326)	(0.01124)	(0.01113)	(0.01113)
λ_i (Presample yogurt purchases)^2	−0.00014	−0.00012	−0.00011	−0.00010	−0.00010
	(0.00002)	(0.00002)	(0.00002)	(0.00002)	(0.00002)
λ_i (Presample Yoplait purchases)	0.05463	0.04687	0.04216	0.04100	0.04129
	(0.01583)	(0.01410)	(0.01319)	(0.01195)	(0.01195)
λ_i (Presample lowfat purchases)	0.04221	0.03549	0.03272	0.03090	0.03136
	(0.01667)	(0.01487)	(0.01322)	(0.01252)	(0.01256)
σ_λ	2.13160	1.76610	1.62990	1.51500	1.51040
	(0.17652)	(0.16045)	(0.28756)	(0.22596)	(0.21967)
Log likelihood	−3958.3624	−3942.3477	−3955.6524	−3943.6655	−3944.0053
Informative advertising effect—1 week*	0.20384	0.39247	0.17026	0.18841	0.17477
	(0.11962)	(0.17159)	(0.08390)	(0.07528)	(0.06206)
Informative advertising effect—5 weeks*	0.38684	0.43850	0.35706	0.37724	0.34987
	(0.17380)	(0.16671)	(0.18483)	(0.14463)	(0.12727)
Informative advertising effect—20 weeks*	0.69631	0.46575	0.62958	0.57476	0.52950
	(0.23173)	(0.18467)	(0.32990)	(0.17113)	(0.14201)
$[\Sigma]_{11}^{1/2} = (\sigma_i^2 + \sigma_j^2)^{1/2}$*	1.99367	2.43135	2.65073	1.94761	1.91048
	(0.16888)	(1.08192)	(0.99821)	(0.17032)	(0.15449)

* These effects are functions of the estimated parameters, standard errors obtained by delta method.

heterogeneous components of utility that are not realized until after their first 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 0.981. 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.[35]

Of particular interest in comparing the myopic and fully dynamic results are the estimates of δ, 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 advertising affects consumers mainly through the informational structure of the model. My "prestige" coefficient on advertising directly in the utility function, θ_3, 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, I report estimates of ρ, the correlation coefficient associated with Σ_0, rather than β_1. All estimates of ρ 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 "elasticities" are all significant and fairly consistent across the models, the only anomaly being in the myopic, time-trend model where information seems to be conveyed quicker, but asymptotes to a lesser magnitude. The overall advertising elasticities generated by the models including time trends are approximately 0.15, very similar to those found in Ackerberg (2001).[36]

[35] In the dynamic models including a time trend, reasonable values (above 0.9) 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. I therefore fix the discount factor at its point estimate from the dynamic model without the time trend.

[36] 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 $\frac{\epsilon_A}{-\epsilon_P}$). 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 deficient.

TABLE 3

GOODNESS OF FIT

Number of Prior Purchases	Data	Model
0	0.00286 (0.00016)	0.00280
1	0.01747 (0.00144)	0.01525
2	0.03501 (0.00375)	0.04179
3	0.04530 (0.00591)	0.06402
4	0.08566 (0.01207)	0.09323
5	0.07368 (0.01198)	0.11427
6	0.12385 (0.02231)	0.13107
7	0.16547 (0.03151)	0.15450
8	0.13291 (0.02700)	0.17067
9	0.28378 (0.05240)	0.18147
10	0.20879 (0.04260)	0.20961

NOTE: Standard deviations of data probabilities in parentheses.

As a check of the goodness of fit in my model, Table 3 compares conditional choice probabilities predicted by the model (the last column of Table 2) to conditional choice probabilities in the data. I 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 examines three perturbations of the myopic model (these perturbations make the dynamic model infeasible to estimate). In the first column I allow a more flexible persuasive effect of advertising, including dummies for m_{it}^a lying in different regions. The estimates are imprecise. I also tried using nonlinear functional forms for m_{it}^a, but did not find much. The second column includes a random coefficient on persuasive advertising, allowing heterogeneity in prestige or image effects across the population. Although the estimated standard deviation of the random coefficient is economically large, it is insignificant even when simulation error is neglected.[37] The mean prestige effect stays at essentially zero. I hesitate

[37] Because there are three unobservables to integrate out over in this model, I switch from discretizing the integrals to using Monte Carlo with crude importance sampling. I also use this technique in the next set of 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.

TABLE 4

ADDITIONAL ESTIMATES

Parameter	Myopic w/Flexible Prestige Advertising	Myopic w/Random Coefficient on Prestige Advertising	Myopic w/Multi-Period Learning ($[\Phi]_{11} = 1$)
θ_1 (Price)	−5.52580	−5.62280	−5.72480
	(0.33081)	(0.32840)	(0.33384)
θ_2 (Store coupon)	3.12930	3.06820	3.11890
	(0.81223)	(0.82226)	(0.81513)
θ_3 (Prestige advertising)		−0.00803	−0.03285
		(0.05651)	(0.04685)
θ_4 (Competitor's price)	−0.69068	−0.71465	−0.72726
	(0.22560)	(0.22015)	(0.22448)
θ_5 (Time trend on	1.23630	1.07650	1.28130
outside alternative)	(0.16740)	(0.17547)	(0.19354)
σ_i (Variance of δ_i	1.76940	1.48520	2.30260
around δ)	(0.14222)	(0.14672)	(2.59952)
σ_j (Consumer's percieved	1.45280	1.70980	4.90710
variance of δ)	(0.14235)	(1.38625)	(1.64029)
ρ (Correlation coefficient	0.62582	0.74675	0.79062
of Σ_0—informative advertising)	(0.04930)	(0.27701)	(0.34281)
δ (Mean experience quality	−0.75524	−0.84418	1.72740
of Yoplait 150)	(0.09484)	(0.76762)	(0.15174)
Discount factor	0	0	0
λ_i - Constant	−3.43571	−3.22620	−3.01320
	(0.44095)	(0.85789)	(1.30272)
σ_λ	1.76680	1.74530	1.73290
	(0.15519)	(0.11035)	(0.12572)
$I(1 <= m_{it}^a < 2)$	0.15037		
	(0.09829)		
$I(2 <= m_{it}^a < 3)$	0.06883		
	(0.14448)		
$I(3 <= m_{it}^a < 4)$	0.19758		
	(0.15790)		
$I(4 <= m_{it}^a < 5)$	0.23300		
	(0.16749)		
$I(5 <= m_{it}^a < 6)$	−0.04442		
	(0.21307)		
$I(6 <= m_{it}^a)$	−0.32607		
	(0.32196)		
S.D. of random		0.11258	
coefficient on θ_3		(0.07562)	
Log likelihood	−3940.0390	−3939.6724	−3921.7334
Informative advertising	0.34616	0.38097	0.53471
effect—1 week*	(0.03477)	(0.17062)	(0.23470)
Informative advertising	0.38876	0.41891	0.58903
effect—5 weeks*	(0.04632)	(0.18613)	(0.21620)
Informative advertising	0.41421	0.44082	0.62047
effect—20 weeks*	(0.05497)	(0.20523)	(0.23471)
$[\Sigma]_{11}^{1/2} = (\sigma_i^2 + \sigma_j^2)^{1/2}$	2.28941	2.26478	2.88182
	(0.14028)	(1.04530)	(2.08098)

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 2. Not all λ_i terms shown.

* These effects are functions of the estimated parameters, standard errors obtained by delta method.

to make any strong conclusions about these results because the random coefficient may be picking up measurement error in my advertising variable. The third column relaxes my assumption of a one-period learning process, fixing σ_v^2 arbitrarily at 1. This results in a path of posterior experience quality variances (in number of prior purchases) of 2.882, 0.742, 0.426. 0.299, 0.230, 0.187,.... Again, we obtain an insignificant prestige effect and a positive, significant informative effect. In summary, my 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.[38]

There is a relatively clear cost–benefit trade-off with my signaling effect of advertising: costs the cost of resources devoted to advertising, benefits the information conveyed to consumers by the advertising (I 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., disprestige) from the utility derived from consuming the outside alternative. Thus my 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 2) 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 that I 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 I 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

[38] It is also dependent on numerous and likely simplistic firm behavioral assumptions and equilibrium extrapolations that will follow. As a result I do not take this to be the final conclusion on the welfare effects of advertising in this market. However, I 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.

period."[39] Another problem is that we do not observe a single price, but rather a distribution of prices. Again in order to simplify things, I assume that the firm chooses a mean price with the price distribution around that mean price fixed (as observed in my sample—perhaps due to retailer behavior).

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

$$\Pi(\bar{p}, a) = TR(\bar{p}, a) - q(\bar{p}, a) \cdot mc - c_a a$$

where $q(\bar{p}, a)$ and $TR(\bar{p}, a)$ 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. $TR(\bar{p}, a)$ and $q(\bar{p}, a)$ are quantities that I can simulate using my 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 of prices. Differentiating with respect to both $\bar{p}$ and a and manipulating the two F.O.C.'s gives

$$\begin{aligned} mc &= \frac{\frac{\partial TR(\bar{p},a)}{\partial \bar{p}}}{\frac{\partial q(\bar{p},a)}{\partial \bar{p}}} \\ c_a &= \frac{\partial TR(\bar{p}, a)}{\partial a} - \frac{\partial q(\bar{p}, a)}{\partial a} mc \end{aligned}$$

Using my estimates, I simulate these four derivatives and solve these equations, obtaining a marginal cost of Yoplait 150 of \$0.422 ($\bar{p} = \$0.653$) and a cost per unit of advertising intensity of \$3732.[40]

With these costs in hand, we can analyze the welfare effects of a ban on advertising. I first eliminate the covariance term in consumers' priors (i.e., set $\beta_0 = \beta_1 = 0$). 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 $\beta_0 = \beta_1 = 0$, I obtain a new demand system ($TR(\cdot)$ and $q(\cdot)$ functions), which is used to numerically find the new $\bar{p}$ that solves the above first order conditions (unfortunately, I cannot optimally adjust competitors' prices because I have no model of demand for

[39] I take this introductory period to be the length of my 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.

[40] This advertising cost is given my simulated data set of 1.2 million consumers. Given I assume constant marginal costs of production, I 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.)

TABLE 5
YOPLAIT 150 WELFARE RESULTS

Variable	Estimated Equilibrium	Ad Ban Adjusting Price Optimally	Ad Ban w/o Adjusting Price
Mean price	0.6527	0.6458	0.6527
Marginal cost	0.4221	0.4221	0.4221
Total sales	236,371.00	225,498.00	218,377.00
Total revenue	148,360.93	140,129.22	137,047.13
Production costs	99,762.744	95,173.686	92,168.197
Advertising costs	9329.996	0	0
Total costs	109,092.74	95,173.686	92,168.197
Profits	39,268.186	44,955.529	44,878.933
Compensating variation	45,541.838	46,554.529	44,962.923
Total welfare	84,810.024	91,510.058	89,841.856

competing products). We find a new profit 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 ($\delta = 0.899 > 0$). 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 exhibit some welfare measures of the two equilibria. Both firm profits *and* compensating variation (CV) increase under the advertising ban. Yoplait does make less variable profits, 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 first column to here measures the pure 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 δ. The linear signaling equilibrium I have estimated is an equilibrium over *many* possible new products with many different experience qualities δ. The benefits of signaling information should differ for different δ's. Intuitively, the benefit should be greatest for products with experience qualities far away from consumers' initial priors on δ, 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 my consumer model—it is the consumers' initial prior on δ (which I have assumed normal, normalized its mean to 0, and estimated the standard deviation $\sigma_j = 0.593$).

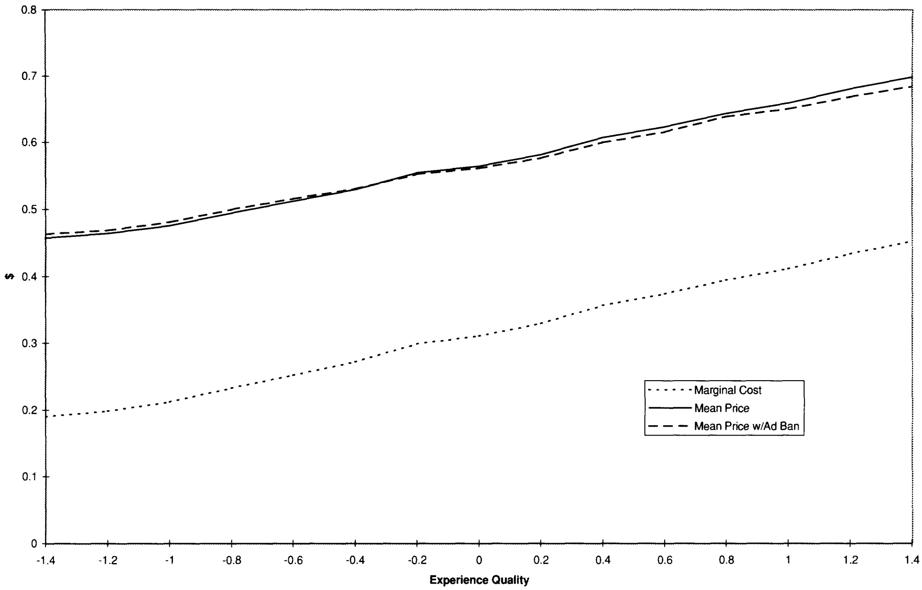

FIGURE 1

PRICES AND MARGINAL COSTS

Unfortunately, I have no direct data on these other "products" with different δ's. However, my estimated model allows I to compute demand for alternate experience qualities $(TR(\bar{p}, a, \delta), q(\bar{p}, a, \delta))$. In addition, although I do not know marginal costs and optimal mean prices for alternative experience qualities ($mc(\delta)$ and $\bar{p}(\delta)$), my estimated signaling equilibrium equation *does* tell us how much they should optimally advertise ($a(\delta) = \beta_0 + \beta_1\delta$). Therefore, as we have two first-order conditions and just two unknowns (mc and $\bar{p}$), I can numerically solve out these unknowns for each δ. Knowing $mc(\delta)$, we can then consider an advertising ban, inverting out prices under the ban. Figure 1 plots the results: $\bar{p}(\delta)$, $mc(\delta)$, and $\bar{p}^{noad}(\delta)$—all increase in experience quality.[41] Of particular note is the result that under the ban, prices become more equalized. Due to the lessened information, higher quality products are less able to price their quality while lower quality products can extract more "dis-information" rents.

Figure 2 plots CV under the two regimes, as well as for the case where the firms 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

[41] The fact that prices vary over experience quality raises the question of why consumers cannot 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 (1986), firms in equilibrium need to set both prices and advertising levels appropriately for a credible signal. A more logistical problem is that my linear advertising equilibrium equation indicates low δ_js (those more than 1.54 standard deviations below the mean) should advertise negative amounts. In simulating welfare for these qualities, I assume that no money is spent on advertising but that consumers get the correct signals anyway, again likely biasing my results slightly in favor of advertising.

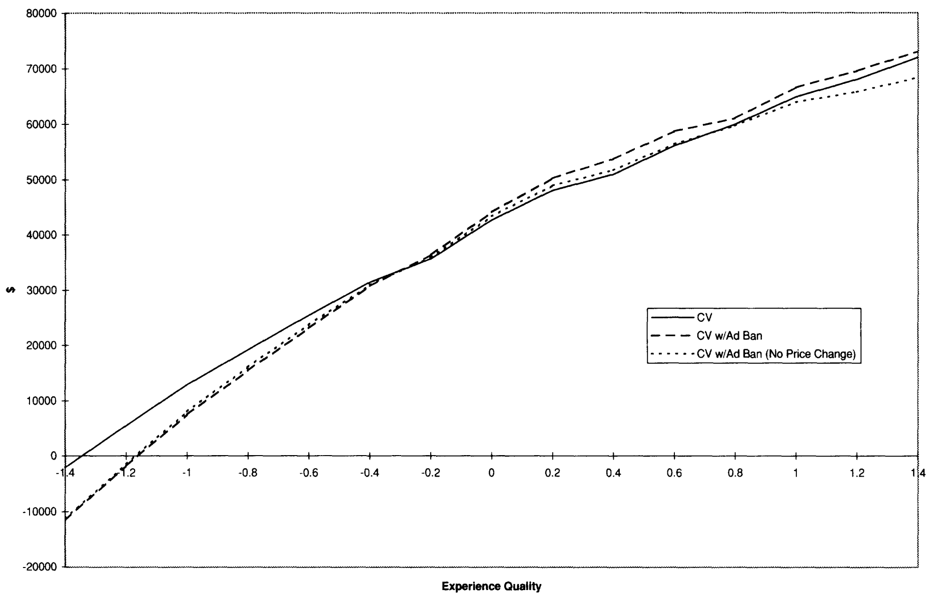

FIGURE 2

COMPENSATING VARIATION

their prior mean the more consumers benefit from the advertising information (Figure 3). 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.[42] 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 $\delta = 1.4$ they may not.)

Figures 4–7 indicate firm 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 firms is such that firms 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 firm 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.[43]

[42] It appears that the lack of symmetry of the "value of information" function (the minimum being at experience quality 0.2) arises from a somewhat inconsistent treatment of the discount factor. In adding my welfare measures over periods I use discount factor above the estimated 0.98 (otherwise things die out very quickly). This means that consumers are actually not behaving to exactly maximize my CV measure. This lower discount factor implies less experimentation behavior, and therefore from my social planner standpoint it is better to fool the consumers into a bit more experimentation.

[43] 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. I suspect that this slope might disappear or change sign if one allowed firms some flexibility in changing prices over time.

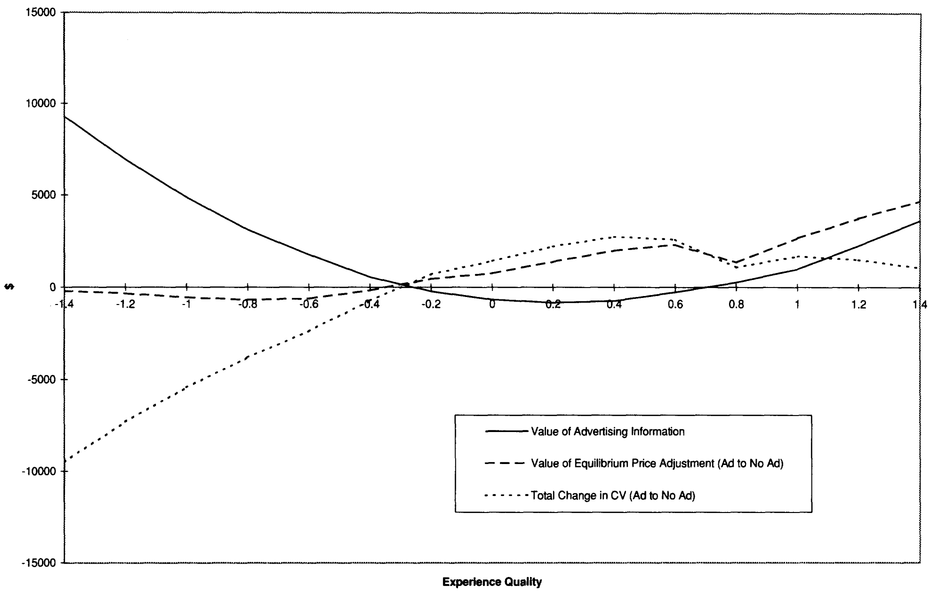

FIGURE 3

DECOMPOSITION OF EFFECTS OF BAN ON CONSUMERS

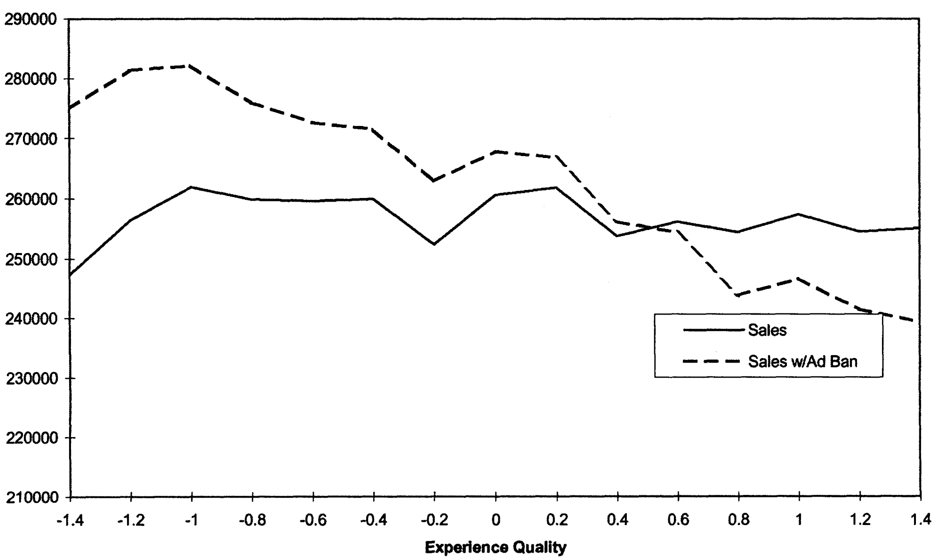

FIGURE 4

SALES

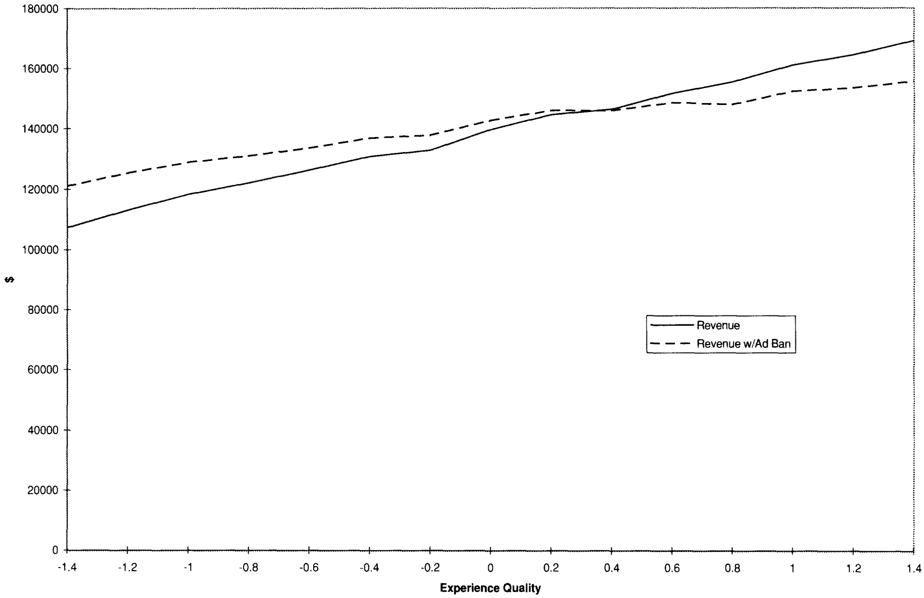

FIGURE 5

REVENUE

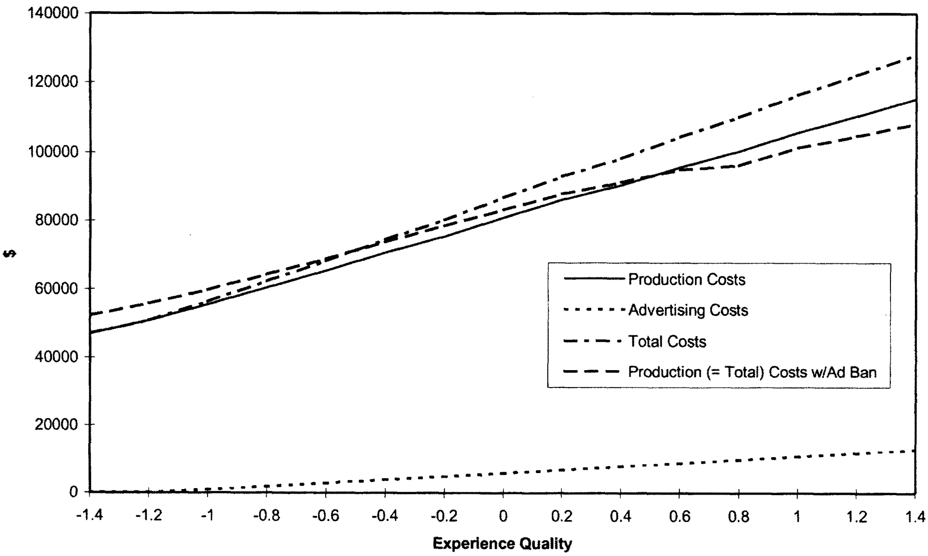

FIGURE 6

COSTS

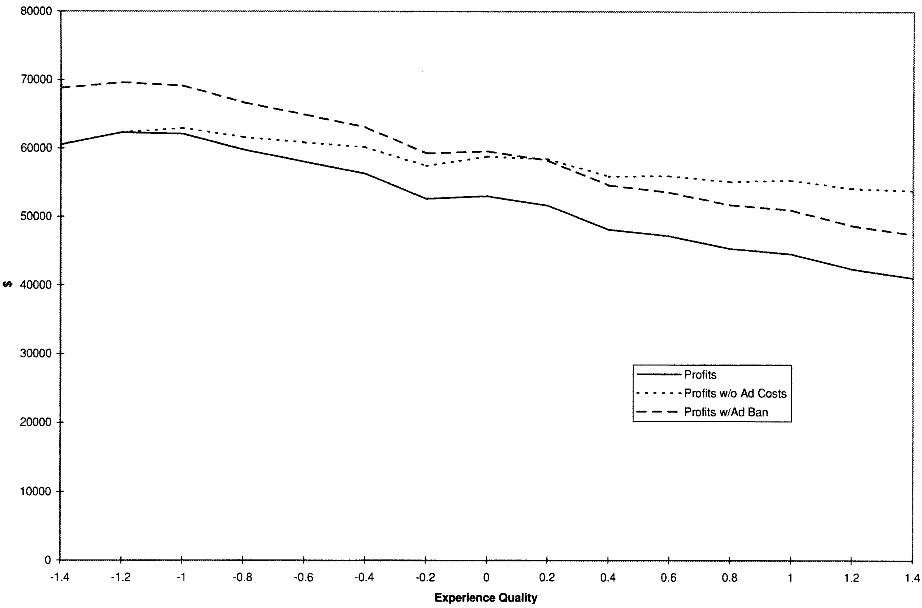

FIGURE 7

PROFITS

More appealing is the fact that the advertising ban slightly accentuates the slope of the profit curve. This suggests that the ability to advertise would increase incentives (or at least decreases disincentives!) to invest in experience quality in a model where such choices were endogenized. Figure 8 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 δ_i distribution divided by total revenues (similarly weighted) implies a welfare 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, I feel my structural estimation results are thought provoking. I present a model in which I explicitly include two effects of advertising: an informative effect that enters the information structure of my 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

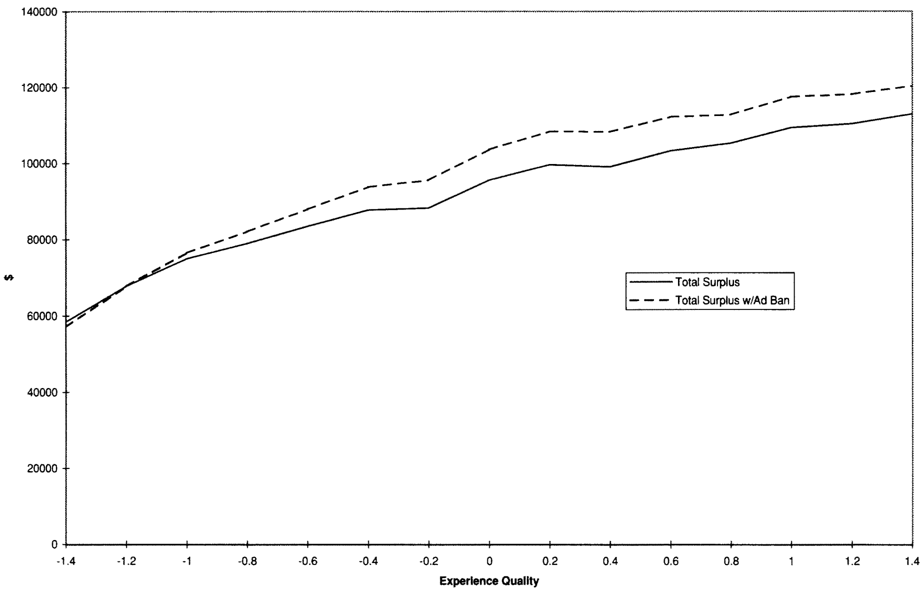

FIGURE 8

TOTAL SURPLUS

learning. I 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. I 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. I take a brief stab at such questions in my welfare examination, albeit in a somewhat unsatisfying way as the analysis rests on some very simple and strong assumptions on firm behavior and my 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. I 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.

REFERENCES

ACKERBERG, D., "Advertising, Information, and Consumer Choice in Experience Good Markets," Ph.D. Dissertation, Yale University, 1997.

——, "Empirically Distinguishing Informative and Prestige Effects of Advertising," *RAND Journal of Economics* 32 (2001), 100–18.

Advertising Age (Chicago: Crane Communications, Serial).

BECKER, G. S., AND K. M. MURPHY, "A Simple Theory of Advertising as a Good or Bad," *Quarterly Journal of Economics* 108 (1993), 942–64.

BENHAM, L., "The Effect of Advertising on the Price of Eyeglasses," *Journal of Law and Economics* 15 (1972), 337–52.

BOYER, K. D., "Informative and Goodwill Advertising," *The Review of Economics and Statistics* 56 (1974), 541–48.

BUTTERS, G., "Equilibrium Distributions of Sales and Advertising Prices," *The Review of Economic Studies* 44 (1977), 465–91.

DEGROOT, M., *Optimal Statistical Decisions* (New York: McGraw Hill, 1970).

DEIGHTON, J., C. HENDERSON, AND S. NESLIN, "The Effects of Advertising on Brand Switching and Repeat Purchasing," *Journal of Marketing Research* 31 (1994), 28–41.

ECKSTEIN, Z., AND K. WOLPIN, "The Specification and Estimation of Dynamic Stochastic Discrete Choice Models," *The Journal of Human Resources* 24 (1991), 563–97.

——, D. HORSKY, AND Y. RABAN, "An Empirical Dynamic Model of Optimal Brand Choice," Mimeo, Tel-Aviv University, 1988.

ERDEM, T., AND M. KEANE, "Decision-Making under Uncertainty: Capturing Dynamic Brand Choices in Turbulent Consumer Goods Markets," *Marketing Science* 15 (1996), 1–20.

GALBRAITH, J. K., *The Affluent Society* (Boston: Houghton-Mifflin, 1958).

GONUL, F., AND K. SRINIVASAN, "Modeling Multiple Sources of Heterogeneity in Multinomial Logit Models: Methodological and Managerial Issues," *Marketing Science* 12 (1993), 213–30.

GROSSMAN, G., AND C. SHAPIRO, "Informative Advertising with Differentiated Products," *The Review of Economic Studies* 51 (1984), 63–81.

GUADAGNI, P., AND J. LITTLE, "A Logit Model of Brand Choice Calibrated on Scanner Data," *Marketing Science* 2 (1983), 206–38.

KEANE, M., "A Computationally Practical Simulation Estimator for Panel Data," *Econometrica* 62 (1994), 95–116.

——, AND K. WOLPIN, "The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence," *The Review of Economics and Statistics* 76 (1994), 648–72.

KIHLSTROM R., AND M. RIORDAN, "Advertising as a Signal," *Journal of Political Economy* 92 (1984), 427–50.

LESLIE, P., "Price Discrimination in Broadway Theater," Mimeo, UCLA, 1999.

LUCAS, R., "Econometric Policy Evaluation: A Critique," in K. Brunner and A. Meltzer, eds., *The Phillips Curve and Labor Markets* (Amsterdam: North Holland, 1976).

MARSHALL, A., *Industry and Trade: A Study of Industrial Technique and Business Organization* (London: MacMillan, 1919).

MCCULLOCH, R., AND P. ROSSI, "An Exact Likelihood Analysis of the Multinomial Probit Model," *Journal of Econometrics* 64 (1994), 207–40.

MILGROM, P., AND J. ROBERTS, "Price and Advertising Signals of Product Quality," *Journal of Political Economy* 94 (1986), 796–21.
MILLER, R., "Job Matching and Occupational Choice," *Journal of Political Economy* 92 (1984), 1086–20.
MILYO, J., AND J. WALDFOGEL, "The Effects of Price Advertising on Prices: Evidence in the Wake of 44 Liquormart," *American Economic Review* 89 (1999), 1081–96.
NELSON, P., "Advertising as Information," *Journal of Political Economy* 82 (1974), 729–53.
PAKES, A., "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks," *Econometrica* 54 (1986), 755–85.
PEDRICK, J., AND F. ZUFRYDEN, "Evaluating the Impact of Advertising Media Plans: A Model of Consumer Purchase Dynamics Using Single Source Data," *Marketing Science* 10 (1991), 111–29.
RESNICK, A., AND B. STERN, "An Analysis of the Information Content in Television Advertising," *Journal of Marketing* 41 (1977), 50–53.
RUSSELL, G., AND W. KAMAKURA, "Understanding Brand Competition Using Micro and Macro Scanner Data," *Journal of Marketing Research* 31 (1994), 289–307.
RUST, J., "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica* 55 (1987), 999–35.
STIGLER, G. L., "The Economics of Information," *Journal of Political Economy* 71 (1961), 213–25.
STIGLER, G., AND G. BECKER, "De Gustibus Non Est Disputandum," *American Economic Review* 67 (1977), 471–99.
TELSER, L. G., "Advertising and Competition," *Journal of Political Economy* 72 (1964), 537–62.
TIROLE, J., *The Theory of Industrial Organization*, (Cambridge, MA: MIT Press, 1988).
WOLPIN, K., "An Estimable Dynamic Stochastic Model of Fertility and Child Mortality," *Journal of Political Economy* 92 (1984), 852–74.