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

Daniel A. Ackerberg*

This version: June 28, 1998

Abstract

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

JEL Classifications: D12, M37, D83

* Economics Dept., Boston University, Boston, MA 02115 (ackerber@bu.edu). This paper is a revised version of the second and third chapters of my doctoral dissertation at Yale University. Many thanks to my advisors: Steve Berry and Ariel Pakes, as well as Lanier Benkard, Russell Cooper, Gautam Gowrisankaran, Sam Kortum, Mike Riordan, John Rust, Roni Shachar, and many seminar participants, including most recently those at the NBER 1997 Winter IO meetings, 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 in this paper are my own.

1. Introduction

Theoretical work in economics has long been concerned with different influences of advertising on consumer behavior. Marshall (1919) praised "constructive" advertising, which he 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 engaging in such behavior. More recently, economists have developed formal models of advertising's possible effects. Stigler (1961), Butters (1977), and Grossman and Shapiro (1984) examine models in which firms send consumers advertising messages to explicitly inform them of their brand's existence or observable characteristics. In contrast to this *explicit* provision of information, Nelson (1974), Schmalensee (1977), Kihlstrom and Riordan (1984), and Milgrom and Roberts (1986) analyze models in which firms producing non-durable experience goods use advertising to *implicitly* signal information on their brand's *experience* characteristics (e.g. unobserved quality or taste). In these equilibria, brands with higher unobserved quality advertise more and consumers rightfully interpret these high advertising levels as signaling information on this higher quality.

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

One finding of this theoretical literature is that the way (or ways) in which advertising affects consumers is an important component of the functioning of a market. Advertising that provides information on a brand's 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[1]. Unfortunately, the theoretical literature cannot tell us which of these effects exist or predominate in a particular market. In certain markets, casual empiricism may suggest an answer[2]. On the other hand, we feel that there is a wide range of markets, some in which advertising expenditures reach more than 10% of revenues, where the answer is not clear. Past formal empirical literature addressing this question has suffered from a variety of problems. Telser (1964) and Boyer (1974) correlate advertising levels and measures of profitability at the industry level. Though interesting, their identifying hypothesis, that informative effects should reduce entry barriers and profitability while non-informative effects should raise them, suffers from acknowledged endogeneity problems[3]. Benham (1972) provides a fascinating study of the consequences of removing legal restrictions on eyeglass advertising, but this relies on a unique natural experiment. Nelson (1974) includes some interesting empirical work that seems to suggest the existence of signaling information in advertising, but his methods cannot formally measure or separate different effects. Resnik and Stern (1978) examine actual advertisements to assess informational content. Unfortunately, information that a product exists or implicit signaling information need not be embodied in explicit verbal or visual content.

This study follows Ackerberg (1996) in capitalizing on recently collected consumer level panel data

[1]One example is entry. If advertising purely provides information, ability to advertise may decrease "informational" barriers to entry in an industry (see e.g Tirole (1988) pg. 289). On the other hand, prestige effects that persist over time might increase barriers to entry , perhaps also creating product differentiation and market power. Of course, we must also realize that advertising content is endogenous and chosen by firms. But the way in which advertising works in a particular market may reflect how "prestige" prone a product is, as well as the extent to which imperfect information exists in a market.

[2]Consider the Coke and Pepsi example above. At the opposite extreme, consider classified ads, which clearly provide a great deal of product information.

[3]For example, one might expect informative advertising to be more common in markets with higher existing "informative" barriers to entry.

to distinguish and measure different effects of advertising. We have data following consumers' grocery 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 (on either product existence, search, or experience characteristics), whether these advertisements generated Becker-like prestige or image effects, or whether there was some combination of both types of effects. Ackerberg (1996) addresses this question using a reduced form empirical approach, looking for a differential effect of these advertisements on experienced and inexperienced consumers of the brand (experienced consumers being those who have tried the brand at some point in the past). Since experienced consumers presumably already know of the brand's existence and its observable and unobservable characteristics, he argues that they should not be affected by exposures to informative advertising[4]. On the other hand, he hypothesizes that Becker-like prestige or image effects of advertising should generally affect both inexperienced and experienced users of the brand[5]. Simple reduced form discrete choice models indicate that, all else equal, the advertisements did affect consumers who had never experienced the brand of yogurt before but did not affect experienced consumers. He concludes that the data are consistent with these particular advertisements affecting consumers primarily by providing information.

The present study applies a similar empirical identification argument from a more structural perspective. To more rigorously examine these informational arguments, we formally model consumer information, introducing a model of consumer behavior that explicitly includes both informative and prestige effects of advertising. We proceed with an introduction to and simple example of our model,

[4]One of the noted exceptions to this argument is if advertising provides information on changing search characteristics, e.g. price. Price information, however, is not typically mentioned in the advertisements like those considered here, i.e. national television advertisements for non-durables. Also noted is the possibility that experience characteristics are not learned perfectly with one consumption experience, in which case signaling information in advertising could affect experienced consumers. See Ackerberg (1996) for other exceptions and more discussion.

[5]The idea here is that if, for example, a consumer obtains an extra z utils from consuming a product that is associated (by advertising) with a particular image, seeing such an ad will increase the utility he obtains from consuming the product by z regardless of whether he has purchased the product 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 such effects. On the other hand, a key to empirically distinguishing these prestige or image effects from informative effects is the assumption that they do not somehow interact in the utility function with measures of 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*. In this case prestige or image advertising would also relatively affect inexperienced consumers. Again, Coke and Pepsi may provide evidence that these effects can exist (without such an interaction), as these ads must be affecting experienced consumers.

then detail its important empirical advantages over the afore-mentioned reduced form approach.

The model which we present and estimate is similar to Eckstein, Horsky, and Raban's (1989) dynamic learning model of experience goods with the addition of these two effects of advertising. In each period of our model, dynamically optimizing consumers choose between differentiated brands of a non-durable, experience good. Consumers start the model with imperfect information on a brand's characteristics. They learn about these characteristic *both* through consumption of the brand *and* through informative advertising. Our Becker-like prestige or image effect of advertising enters directly in the utility function, influencing utility independently of beliefs on inherent product characteristics[6]. A related model is estimated in Erdem and Keane (1996). They also extend the model of Eckstein, et. al. to include informative advertising, but are primarily concerned with the demand implications of this single effect and do not try to distinguish different effects of advertising[7].

The most basic representation of the important empirical components of our model is as follows. Consider a consumer who purchases a brand if the utility he expects to obtain from consuming the brand is greater than some threshold k, i.e. iff

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

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 level and/or content. The expectation is over δ as the consumer doesn't necessarily know all the brand's characteristics perfectly. Note that a enters in two places into this expected utility. First, it directly enters the utility function. This is our prestige or image effect of advertising - advertising influencing

[6]We 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. We give the consumer more credit than that.

[7]There are a number of other significant differences between the two empirical models. One is in the extent to which consumer heterogeneity is allowed. Our model focuses on one particular brand and allows consumers to differ in both initial perceptions of the brand and final (post-information) perceptions of the brand. In other words, some consumers learn that they really enjoy the yogurt, while others find out that they do not. In Erdem and Keane, there is no heterogeneity in what is learned. Consumers all converge to the same belief about the utility they will obtain from a brand. On the other hand, they are able to examine learning and advertising for multiple (8) brands. Also, they examine laundry detergent, where idiosyncratic tastes may be less of a factor than in food products. The models also differ in the way that informative advertising is modelled (see below) and in the policy analysis that is performed. They examine and evaluate alternative firm advertising strategies while we measure the value of information contained in advertising.

utility *given* inherent product characteristics. Secondly, the expectation over δ is conditioned on a. This is our informative effect of advertising - we allow advertising to "tell" the consumer something about the brand's characteristics δ[8]. As consumption of the brand also provides information to the consumer on the brand's characteristics, we obtain the result that informative advertising impacts the expected utility of inexperienced consumers more than that of experienced consumers. In the case where consumption of one unit provides *perfect* information on δ, informative advertising does not affect the expected utility or behavior of experienced users at all. On the other hand, our prestige or image effect of advertising affects utility regardless of whether a consumer is experienced or not. This distinction is what separately identifies these two effects of advertising in our empirical work, similar to but in a more structural and formal fashion than the reduced form models above.

Formalizing this model involves specifying the process through which informative advertising affects a consumer's information set. As noted above, there are a number of different types of information advertising can provide: explicit information on product existence or observable characteristics, or signaling information on experience characteristics. It would be optimal to write down and estimate a consumer model including all these possible informative effects. Unfortunately, such a model would likely be computationally intractable, and more importantly, these separate informative effects would be hard, if not impossible, to empirically distinguish given our dataset. We therefore choose just one of these informative effects to include in our structural model, a signaling effect of advertising[9]. Reasons for this particular 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 very casual empirical evidence from Ackerberg (1996), and (3) convenience and flexibility in computation and estimation[10]. Given the necessity of making such a choice, it is *very important* to note that this

[8] An intuitive way of thinking about these two effects within the framework of Lancaster's (1971) characteristics-based product differentiation is that our informative advertising tells the consumer where a product lies in characteristic space; our prestige or image advertising involves advertising actually constituting a dimension in characteristic space.

[9] The effect of advertising modeled in Erdem and Keane (1996) is a different type of informative advertising, one where each advertisement a consumer sees tells him some explicit information about the product.

[10] Regarding (2), one hypothesis is that information on existence or explicit characteristic information should depend more on the absolute number of advertising exposures a consumers sees than advertising "intensity" (i.e. advertisements seen per hours of TV watched) . On the other hand, signaling information might depend more on advertising intensities, the consumer wants to know how much a brand is spending on advertising. In the reduced form models of Ackerberg (1996) advertising intensities fit the data better than the absolute number of advertising exposures. This could be suggestive of signaling (on the other hand, it also might indicate that the intensity variables have less measurement error

empirical work *does not* take a stand on which types of informative effects of advertising are actually occurring in our market. However, we believe that these different informative effects of advertising should in some sense be observationally equivalent in our data: all tend to affect inexperienced rather than experienced consumers of the brand. As a result, we feel that our empirical analysis and conclusions regarding significance or insignificance of our informative and prestige effects would not substantially change if we had instead modeled one of the other informative effects of advertising. In summary, we interpret a statistically significant signaling effect of advertising *not* as empirical support for signaling *per se*, but as support for the more general hypothesis that advertising is providing some kind of product information to consumers.

There are important advantages of this structural approach to distinguishing different effects of advertising as compared to the reduced form models of Ackerberg (1996)[11]. 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 our structural learning model, which does accommodate such dispersion by allowing heterogeneous consumer tastes for the brand that are not realized (learned) by a consumer until *after* having experienced the brand[12]. Not only will ignorance of this dispersion be inefficient, but it can potentially generate

than the absolute ones). Regarding (3), we also speculate that prestige effects might depend on advertising intensities rather than a consumer's absolute number of exposures (prestige may be generated by the amount of advertising a brand does, image effects may depend on how intensely a product is associated with an image). Thus, with a signaling effect, we only have to keep track of one advertising variable per consumer (intensity) rather than two (both intensity and number of exposures), and identification comes from the more robust implications of the model rather than by definition of two different advertising measures.

[11] There are also notable disadvantages, including 1) more structural assumptions, including the restriction to only explicitly including one informative effect of advertising, and 2) increased computational complexity, which prevents as exploratory an analysis as one might like.

[12] In the reduced form models, adding an unobserved interaction term (e.g. a persistent random coefficient) on a dummy variable "prior experience" might be able to partially replicate this dispersion. However, such models begin to look a lot like the myopic structural models used in this paper.

spurious results[13]. These types of issues illuminate the need to consider structural models in future 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 the reduced form analysis. 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. In our case we ban advertising and are able to adjust consumer behavior appropriately so that the resulting zero advertising levels are not interpreted as a "bad" signal. We stress both here and later that, unlike our main empirical conclusions, the welfare analysis we perform is probably highly dependent on our choice to model informative advertising as a signaling effect. Though this limits the applicability of the welfare results, we feel that it is still an interesting and enlightening exercise.

Estimates of our structural learning model support two main conclusions. First, we can easily reject the hypothesis of perfect information. The data suggest that consumers do learn from their consumption experiences with the brand. Second, we find a strong, positive informative effect of advertising and an economically and statistically insignificant prestige effect of advertising. This supports the reduced form conclusion that the advertisements in this data primarily affected consumers through the provision of information. Again, we stress that since we include only one informative effect, we are prevented from drawing any conclusions about the nature of informative effect, i.e. whether it is in fact signaling information or perhaps information on product existence or observable characteristics. Under the strong assumption that it is in fact signaling information, our policy analysis

[13] As Ackerberg (1996) notes, their argument regarding different empirical implications of informative and prestige advertising is made *conditional* on a consumer's expected utility (EU) from consuming the brand. In other words, empirically one wants to compare the effects of a brand's advertising on two consumers with the *same* EU from consumption, but *different* levels of experience with the brand (e.g. one experienced, the other inexperienced). The argument needs to be conditional because of potential correlations in the data between prior experience and current expected utility from consumption (see Ackerberg (1996) for a simple example). Though such positive correlation between prior experience and expected utility is likely adequately conditioned on in the reduced form models, the increased *dispersion* in behavior (EU) mentioned in the text above is not. For example, consider a situation where both experienced and inexperienced consumers have EU's centered at zero (assume consumers purchase if $EU > 0$), but experienced consumers EU's have more dispersion (a higher variance) (see Figure 1). In this case, a burst of prestige advertising that shifts *all* consumers EU's up by a certain amount will induce a higher proportion of inexperienced consumers than experienced consumers to purchase. Without conditioning on this increased dispersion, one would incorrectly conclude that this advertising relatively affects inexperienced consumers.

indicates that the value of this information to consumers is significantly less than the resources spent on advertising. This at least suggests that advertising signaling may be a very inefficient way of transferring information. Section 2 introduces our general model of consumer behavior and Section 3 describes the data used in this study. Section 4 details our empirical specification and presents our results. In Section 5 we perform our welfare experiments and Section 6 concludes.

2. The Model

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

After observing prices and advertising intensities in a given period, the consumer decides whether to purchase one unit of one of the brands or nothing. 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 $c_t \in \{0, ..., J\}$ is the consumer's choice at t (0 represents no purchase) and β is the per-period discount factor[14].

As is now relatively common in the empirical analysis of differentiated products, we take a Lancasterian, characteristics-based approach to consumer theory, assuming the utility a consumer derives from a brand is a function of the brand's characteristics and the consumer's tastes for these character-

[14]Note that we consider an infinite horizon problem. As consumers have finite lives, there is obviously some finite horizon, but because the time frame of the empirical work will be a consumer shopping trip, the number of periods will be very large and approach the infinite horizon solution. Also, note that because the consumer's information may change through time, the consumer maximization problem is over a sequence of choice *functions* mapping future realizations of information sets into choices.

istics. Specifically, we assume the utility consumer i obtains from purchase and consumption of brand j in period t is:

$$U_{ijt} = U(p_{ijt}, X_j, y_i, \epsilon_{ijt}, m^a_{ijt}, \delta_{ijt+1}) \tag{2.2}$$

where X_j contains *observable* characteristics of brand j, and y_i are consumer i's tastes for these characteristics. ϵ_{ijt} represents idiosyncratic, time-varying shocks to the utility the consumer derives from consuming brand j that are *known* prior to the purchase decision. Though we defer its formal definition until later, m^a_{ijt} is a measure of what consumer i currently knows about how much brand j is advertising. Its entry into the utility function will represent our image or prestige effect of advertising.

The term δ_{ijt+1}, which we call "experience utility", captures the experience nature of the good. It is a scalar measure of the utility that consumer i derives from brand characteristics that are not directly observable to him (i.e. experience characteristics). It is dated $t+1$ because in contrast to the other elements of the utility function, it is *not* necessarily known to the consumer at the time of purchase. For food products, X_j might contain observable characteristics such as calories or fat content, while δ_{ijt+1} might represent how the brand actually tastes to the consumer (conditional on X_j). Although δ_{ijt+1} is not observable before purchase, it is observed if good j 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_{ijt+1} is constant over time we have a "one-consumption" learning process. In this case, if the consumer purchases and consumes the brand once, he observes δ_{ijt+1} and knows its value for all future t.

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

$$\delta_{ijt+1} = \delta_{ij} + \nu_{ijt+1} \qquad \text{where } \nu_{ijt+1} \sim iid\ N(0, \sigma^2_\nu) \tag{2.3}$$

Although δ_{ijt+1} is realized (observed) by the consumer after consumption, its components, δ_{ij} and ν_{ijt+1}, are never individually observed. δ_{ij} is the mean experience utility consumer i obtains from brand j. ν_{ijt+1} are i.i.d. confounding variables that cannot be distinguished from this mean. In the case of food products, variance in ν_{ijt+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[15]. In contrast to the i.i.d. ν_{ijt+1}, δ_{ij} is persistent over time. It is thus beneficial for the consumer to use information contained in observed δ_{ijt+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_{ij} (and thus $\delta_{ijt+1}\ \forall t$) is learned after one consumption experience. In the non-degenerate case, consumption and subsequent realization of δ_{ijt+1} does not exactly reveal δ_{ij}, but it does provide information about it. This information will be consistently modeled in a Bayesian learning framework.

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

$$a_{ijt} = a_j + \xi_{ijt} \qquad \text{where } \xi_{ijt} \sim iidN(0, \sigma_\xi^2) \tag{2.4}$$

and where a_j is the mean advertising intensity of brand j. Deviations in a_{ijt} around a_j may be caused by variation in consumers' television or reading habits or variation in where or when a brand is advertised[16]. Although consumers do not directly observe a brand's mean advertising intensity a_j, we allow them to be interested in it for two reasons: (1) Possible prestige, image or status effects of advertising where the consumer, all else equal, obtains more utility from consuming a more advertised brand or a brand more associated (through advertising) with a particular image, and/or (2) a belief that firms use a_j to implicitly signal information on the mean experience utility they obtain from the brand δ_{ij}, as would be the case in a Nelson type signaling equilibrium. In either of these cases, an optimizing consumer will use observed a_{ijt}'s to learn about a_j. Note that in this model there is no

[15] In all these cases the important thing is that the ν_{ijt} are indistinguishable from δ_{ij}, 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. Note that one could also add randomness to the utility function that is not known prior to purchase but *is* distinguishable (from δ_{ij} and ν_{ijt}) after consumption. As long as such error is additive and i.i.d., only its expectation has an effect on decision making.

[16] One can easily generalize to having advertising exposures distributed around an individual specific mean (i.e. $a_{ijt} = a_{ij} + \xi_{ijt}$). It is also possible to allow brands' advertising levels to vary randomly over time. However, these levels need not be serially correlated to ensure that per-period deviations from an individual's mean are independently distributed (this is necessary for the Bayesian updating formulas below).

explicit information about the product obtained through advertisements: consumers are assumed to know of the existence of all brands and know all the *observable* characteristics of each brand[17].

We consistently model information provided by the observed a_{ijt}'s and δ_{ijt+1}'s on the relevant unknowns a_j and δ_{ij} as a bivariate Bayesian learning process. In matrix notation, equations (2.3) and (2.4) become:

$$\begin{pmatrix}\delta_{ijt+1}\\ a_{ijt}\end{pmatrix} \sim \text{i.i.d. } N\left(\begin{pmatrix}\delta_{ij}\\ a_j\end{pmatrix}, \Phi_{ij}\right) \qquad \text{where } \Phi_{ij} = \begin{bmatrix}\sigma_\nu^2 & 0\\ 0 & \sigma_\xi^2\end{bmatrix} \tag{2.5}$$

The assumed diagonality of Φ_{ij} implies that there is no correlation between δ_{ijt+1} and a_{ijt} *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_j and δ_{ij}:

$$\text{Initial Prior:} \qquad \begin{pmatrix}\delta_{ij}\\ a_j\end{pmatrix} \sim N\left(\begin{pmatrix}m_{ij0}^{\delta}\\ m_{ij0}^{a}\end{pmatrix}, \Sigma_{ij0}\right) \tag{2.6}$$

generates a learning process in which a consumer's posterior on brand j after a history of observed advertising intensities, $\{a_{ij1}, ..., a_{ijt}\}$, and consumption experiences, $\{\delta_{ij1}, ..., \delta_{ijK_{ijt}}\}$, is given by[18]:

$$\begin{pmatrix}\delta_{ij}\\ a_j\end{pmatrix} \sim N(\ m_{ijt}\ ,\ \Sigma_{ijt}\ ) \tag{2.7}$$

$$\text{where:} \quad m_{ijt} = \binom{m_{ijt}^{\delta}}{m_{ijt}^{a}} \quad = (\Sigma_{ij0}^{-1} + \eta_{ijt}\Phi_{ij}^{-1})^{-1}(\Sigma_{ij0}^{-1}m_{ij0} + \eta_{ijt}\Phi_{ij}^{-1}\overline{z}_{ijt}),$$

[17]As mentioned and discussed in the introduction, complexity and identification issues necessitated the inclusion of only one informative effect of advertising in our model. There are clearly alternative specifications of informative effects. Erdem and Keane's (1996) similar dynamic model has advertising explicitly informing the consumer on characteristics of the brand. In their model, each advertisement a consumer sees gives them an iid draw from a distribution centered at δ_{ij}. Another alternative would be to allow advertising to inform consumers of a product's existence. Modeling this would likely involve introducing an indicator function of product existence knowledge that is influenced by advertising as well as other things. See the introduction for arguments why we chose our particular effect and why we feel that altering this choice would not significantly affect our primary empirical results.

[18]The derivation of the following conjugate result is a fairly simple extension of the derivation for a multivariate normal with *equal* draws in DeGroot. Note that use of these conjugates (and all the (in our case normality) assumptions that go with it) is necessary because it allows us to express posterior distributions with a finite number of parameters. As the consumer's state space will need to include their current posterior, this reduces the state space from an entire function to a simple set of parameters, allowing the dynamic program to be numerically solvable.

$$\Sigma_{ijt} = (\Sigma_{ij0}^{-1} + \eta_{ijt}\Phi_{ij}^{-1})^{-1} \;, \qquad \overline{z}_{ijt} = \begin{pmatrix} \frac{1}{K_{ijt}} \sum_{k=1}^{K_{ijt}} \delta_{ijk} \\ \frac{1}{t} \sum_{\tau=1}^{t} a_{ij\tau} \end{pmatrix} \text{ ,and} \qquad \eta_{ijt} = \begin{bmatrix} K_{ijt} & 0 \\ 0 & t \end{bmatrix}$$

and where K_{ijt} equals the number of consumption experiences consumer i has had with brand j 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.

The consumer's period t posterior mean for brand j, $m_{ijt} = \binom{m_{ijt}^{\delta}}{m_{ijt}^{a}}$, is a matrix weighted average of initial priors and observed realizations of δ_{ijt+1} and a_{ijt}. An important result of Bayesian learning is that these posterior means and variances summarize all the consumer's information on δ_{ij} and a_j. Thus, under the i.i.d. assumption of (2.5), the current posterior $(m_{ijt},\ \Sigma_{ijt})$ is sufficient to define perceived distributions over both future δ_{ijt+1}'s and a_{ijt}'s as well as future posteriors[19].

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_{ij0} is zero, then the updating processes on δ_{ij} and a_j are completely independent. On the other hand, a non-zero covariance term indicates a perception by consumers that δ_{ij} and a_j are correlated. This links the two learning processes - in this case, observed levels of advertising that provide direct information on a_j will, through this covariance term, provide indirect information on δ_{ij}. Such a correlation in priors would arise from a belief that advertising is used by firms to signal information on a brand's experience utility. As an example, suppose there is a Nelson type signaling equilibrium in which firms set brand advertising levels according to:

$$a_j = \beta_0 + \beta_1 \delta_j$$

where δ_j is brand j's mean experience utility level over the population. Then, assuming (1) a normal population distribution of δ_{ij} around δ_j ($\delta_{ij} \sim N(\delta_j, \sigma_i^2)$) and (2) a normal prior on δ_j ($\delta_j \sim N(m_{ij0}^{\delta}, \sigma_j^2)$), a Bayesian consumer's initial prior variance matrix will have covariance element $\beta_1 \sigma_j^2$. This covariance term captures the informative signaling effect of advertising in our model. In

[19] Note that in this formulation Φ_{ij} is assumed known to the consumer. One implication of this is that the posterior variance matrix evolves deterministically in the number of consumption experiences and time. It is possible to derive updating rules in cases where these variances are not known. Unfortunately, this leads to a structure too complicated for empirical work.

the case where it is positive, consumers interpret high levels of advertising as a signal of a higher δ_{ij}[20].

In addition to this informative effect of advertising, we accommodate prestige or image effects of advertising by allowing the consumer's current posterior mean on a_j, m^a_{ijt}, to enter directly into the utility function (2.2). A positive derivative of (2.2) with respect to m^a_{ijt} indicates that, all else equal (in particular expectations over δ_{ijt+1}), the consumer receives more utility from consuming a product that he believes has a higher advertising intensity. Again, the interpretation is as in Becker and Murphy (1993), i.e. advertising itself or images in advertising bestow some "coolness" or "prestige" on the product that is directly valued by consumers[21].

Given the learning process as specified in (2.7), we can move back to the consumer's dynamic choice problem. Because the posterior $(m_{ijt},\ \Sigma_{ijt})$ is sufficient to define all the consumer's current information on δ_{ij} and a_j, the sequential maximization problem of (2.1) can be transformed into the following Bellman's equation:

$$\begin{aligned} V_i(p_{it}, m_{it}, \Sigma_{it}, \epsilon_{it}) = \max_{c_{it}\in\{0,..,J\}} \ & E[U(p_{ic_tt}, X_j, y_i, \epsilon_{ic_tt}, m^a_{ic_tt}, \delta_{ic_tt+1}) \\ & + \beta V_i(p_{it+1}, m_{it+1}, \Sigma_{it+1}, \epsilon_{it+1}) \mid (p_{it}, m_{it}, \Sigma_{it}, \epsilon_{it}), c_{it}] \end{aligned} \tag{2.8}$$

where the state space $(p_{it}, m_{it}, \Sigma_{it}, \epsilon_{it})$ contains prices, current posterior distributions for each of the J brands, and time-varying preference shocks for all J brands[22]. The expectation is over current period

[20] For the most part (particularly in the welfare analysis that follows estimation), we take the view that consumers are rational in their beliefs, i.e. that a positive covariance term in priors is actually generated by a signaling equilibrium, not just spurious beliefs by consumers. A less restrictive, alternative interpretation is that in estimating $\sum_{ij0}$, we are simply estimating consumer beliefs, without assuming anything about where they come from or whether or not they are correct. In this case, a positive covariance term simply means that consumers think more advertised products will be better, perhaps rightly, perhaps wrongly - making the description of this effect as "information" a rather loose one.

[21] As we are essentially putting "the amount the consumer thinks the brand is advertised" in the utility function, the prestige interpretation is a bit more natural than the image interpretation. At first take, one might consider an image effect of advertising more discrete - a consumer either knows or doesn't know that a brand is associated with a particular image (e.g. Michael Jordan). On the other hand, image effects may arise from a consumer wanting (i.e. deriving utility from) *other* people associating them with particular images. As 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 this use of this term. Clearly we are in somewhat murky waters regarding the specification of these image and prestige effects. To partially compensate, in empirical work we try to be as general as possible with the specification. We also consider the possibility that these effects are heterogeneous across the population (i.e. consumers have different valuations of prestige or different attitudes towards particular images). The most critical thing that we are not flexible with is that m^a_{ijt} doesn't somehow interact in the current period utility function with measures of *past* consumption. As mentioned in the introduction and should be clear in a moment, such interactions can eliminate our source of separate identification of the informative and prestige effects of advertising.

[22] Note three strong assumptions of the model: 1) We assume J independent learning processes for the different brands of the product. Such would not be the case if one were learning about preferences for experience characteristics of the

experience utility δ_{ic_tt+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[23]. 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 an optimal choice of brand. Ackerberg (1996b) 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 on a consumer: First, they may provide indirect, signaling information on these 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. The consumer decides between brands based on the information learned from past consumption and advertising while realizing the effect that current decisions will have on future information.

Unfortunately, the above dynamic model is not analytically solvable. However, we have used numeric solution methods to solve and generate predictions of the model. These general predictions are made through characterization of the value function and simulation of the model at reasonable sets of parameters. The range of parameters tried as well as the intuitive nature of the predictions suggests that they should hold at most if not all parameter values, but this is not necessarily the case. Ackerberg (1996b) contains a more thorough comparative static analysis of this model.

One implication of the learning process is that consumers are likely to change their purchasing

---

general product (e.g. if one has never tried yogurt before, consumption of one particular brand of yogurt will surely provide information on what to expect from consuming a different brand of yogurt), 2) The model ignores potential spillovers of information between consumers. Consumers are unable converse with each other about observed advertising levels or experience utilities. If this were the case, it clearly would provide additional information to consumers on a_j and δ_j, and possibly on their own specific δ_{ij}'s (As might be the case if two consumers knew themselves to have similar preferences for experience characteristics or if experience characteristics were easily describable (e.g. That brand is very sweet).), and 3) The model assumes that there is no signaling information in price. In theory, there is no reason why one couldn't add price signaling to the model in an analogous fashion to the way advertising signaling is modeled.

[23] $V_i(\cdot)$ is indexed by i because of consumer specific, time invariant variables such as y_i and Φ_i that are not explicitly included in the state space. It is the *perceived* value because this is what the consumer believes his expected discounted value of future utilities to be (given his information). There is also a "true" value function, the "true" expected discounted value of future utilities, which is additionally a function of the unknown δ_{ij}'s. This true value function captures the extent to which posteriors differ from the true δ_{ij}, and the "mistakes" that might result.

patterns[24] over time as a result of obtaining new information. The parameters of the learning process determine how long these purchasing patterns will be changing. In a one brand, no advertising, model with a one-consumption learning process ($\sigma_\nu^2 = 0, [\Sigma_{ij0}]_{11} > 0$), purchase patterns change after the first purchase but not thereafter. If there is variance in δ_{ijt+1} ($0 < \sigma_\nu^2 < \infty$), purchasing patterns do change after the second or more purchases, the extent and length depending on σ_ν^2 and $[\Sigma_{ij0}]_{11}$. On the other hand, if there is no learning ($\sigma_\nu^2 = \infty$ or $[\Sigma_{ij0}]_{11} = 0$), 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 more the consumer discounts the future, the less likely he is to experiment as the information benefit to future utilities is weighed less. If advertising is more informative (a large covariance term in Σ_{ij0}) or more precise (σ_ξ^2 is smaller), the consumer is also be less likely to experiment, relying on informative advertising for information rather than trying the new product. On the other hand, the quicker the experience learning process (σ_ν^2 smaller) or the more information there is to be learned ($[\Sigma_{ij0}]_{11}$ larger), the more likely the consumer is to experiment with a new product that does not maximize current expected utility.

The most important implication of the model for the current empirical study concerns the two different effects of advertising in the model. Both consumption and informative advertising can provide information to the consumer on the mean experience utility he will obtain from future consumption. However, while consumption provides direct information on this experience utility, advertising only provides indirect information through a consumer's prior beliefs that the two variables are correlated.

[24]Purchasing patterns generally refer to the number of purchases of a brand in a given time interval. By allowing variance in either ϵ_{ijt}'s or prices, consumers in our model exhibit non-degenerate (i.e. not either always purchase or never purchase) purchase patterns even in the case where there is no learning (Learning itself (without variance in either ϵ_{ijt}'s or prices) will generate purchasing pattern changes, but in a very binary way (e.g. without advertising we have a simple optimal stopping time model)).

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. At that point, the consumer will have already learned δ_{ij} and have no use for the information on experience utility contained in advertising. In contrast, our direct, prestige or image effect of advertising affects the expected utility of inexperienced and experienced users of a brand equally. This is the behavioral implication that we take to the data to distinguish between these two effects of advertising.

3. The Data

Consumer level panel data on grocery purchases is used in estimation of the model of section 2. This data, collected by A.C. Nielsen, is commonly referred to as "scanner panel data" because it was recorded by supermarket UPC scanners[25]. 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 (1986-1988). There is also data on weekly prices at each store, so we essentially know prices on each household's shopping trips[26]. In addition to containing this extremely detailed data on household purchases over time, A.C. Nielsen TV meters were used to collect information on household TV advertising exposures for about half the households in the last year of the data. We thus know, along with when and what each household bought, when members of the household were potentially exposed to TV advertisements for each brand.

The publicly available Nielsen data contains data on four product categories: ketchup, laundry

[25]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.

[26]It is important to note that data on prices in a particular store on a given day is not through direct observation, but rather imputed from purchases by other consumers in the sample (see Ackerberg (1996) for more detail on this issue).

detergent, soup, and yogurt. Ackerberg (1996) chose to focus on the yogurt data for reasons that are just as relevant for this study. First, the inability to even parsimoniously include inventory behavior and purchase quantity choice in the model suggests the choice of the least durable of the above products. Second, empirical identification in both models relies on distinguishing experienced from inexperienced users of a brand. This generates a serious initial condition problem unless one has data from a product's initial introduction on the market. The yogurt data includes such a product: Yoplait 150, a lowfat yogurt introduced in April, 1987, about 15 months before the end of the Nielsen data. As computational issues are even more binding here than in that paper, we again focus specifically on Yoplait 150, modeling competing brands in a informationally static and unfortunately sparse framework.

Table 3.1 gives some summary statistics for the data following Yoplait 150's introduction[27]. Comparing advertising shares to market shares suggests that it was, at least initially, a heavily advertised yogurt. The large difference in market shares between markets 1 and 2 may be due to the existence of two, high-share, local brands in market 1 and the significant number of manufacturer coupons that seem to have been available in market 2. We urge the reader to consult Ackerberg (1996, 1996b) 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[28] and advertising[29].

[27]Only households whose television viewing was recorded are included both here and in estimation. We also limit the data to shopping trips in supermarkets (rather than drugstores) and those in which $10 or more was spent.

[28]Briefly, since we only observe manufacturers coupons that are redeemed, this cannot be used as an explanatory variable due to correlation with any unobservables determining purchases. Because of their relative prevalence in market 2, we use a market dummy as a proxy for the "availability of manufacturers coupons" (this is the variable we really want). In contrast, we do know when store coupons (coupons typically distributed in the store) were available (in the data this was only for one week in two stores), so we do include this as an explanatory variable.

[29]A significant problem with advertising is that advertising exposures were only measured in the last year of the Nielsen study. This leaves about three months when Yoplait 150 was on the market but advertising was not measured. We use zero advertising exposures for this period, a potential justification being that for almost three weeks after TV measurement started, there were no Yoplait 150 advertisements observed. We hope this may indicate that Yoplait did not begin advertising the product until this time. In any case, evidence in Ackerberg (1996) suggests that alternative treatments of this time period do not affect the identification of different effects of advertising. Other problems with our advertising variable include unreliability of TV meters (we eliminated consumers with extremely large viewing gaps - an indication that their meter may not have been working), and a problem inherent in data such as this that we are not sure exactly sure who in a household (if anybody, for that matter) watched or paid attention to an advertisement.

4. Estimation

We now move to estimation of our model using this data, starting with a detailed discussion of our empirical specification. We begin with precise formulations of the utility function and the learning process. As the model contains a significant degree of unobserved consumer heterogeneity, we feel it important to emphasize which variables are econometric observables, which are econometric unobservables, and which are assumed constant over consumers and estimated. It is also important to consider what is known to consumers at particular points of time. Because the empirical model is fairly complicated and non-linear, we intuitively discuss how our data should identify the parameters of the learning process, followed by a description of the likelihood function. We then mention issues regarding solution of the dynamic programming problem generated by the model before concluding with a discussion of our estimates. 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, not the expected discounted sum of all future utilities. 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. Note, however, that this rules out the experimentation behavior generated by the dynamic model. The second set of estimates are of the full dynamic problem in which consumers are forward-looking in their behavior. This does require solution of the dynamic programming problem.

4.1. Empirical Specification

In all empirical work, we assume that the time frame of our model is the consumer (i.e. household)-shopping trip. Specifically, we model consumer i's purchase decision on each of their shopping trips t through the 15 months of data. Because of computational issues, we restrict ourselves to modeling the simple binary choice whether or not to purchase Yoplait 150 on each shopping trip. As such, the choice to purchase a different brand of yogurt is included in the "outside alternative" along with the decision not to buy any brand of yogurt[30]. We assume the following parametric specification of our consumers' single period utility functions:

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

The variables p_{it}, sc_{it}, and p^{oth}_{it} measure price of Yoplait 150, value of a possible store coupon available for Yoplait 150, and a scalar measure of other yogurts' prices[31] 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.

The time-invariant λ_i represents consumer i's individual-specific preferences for the *observable* characteristics of Yoplait 150. λ_i is thus *known* to consumer i for all t. Econometrically, we model λ_i as a linear combination of observable consumer characteristics plus a normally distributed random variable with variance σ^2_λ. These observable consumer characteristics (y_i) include 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[32]. 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.

ϵ_{i1t} and ϵ_{i2t} capture time-varying shocks to the utility derived from each alternative. These are also assumed *known* to the consumer at the time of purchase. To ease computation in both the dynamic

[30] In more preliminary specifications we compared these 2 choice models to 3 choice models (with the choices: Yoplait 150, a different brand of yogurt, or no yogurt) and obtained very similar results. Note that we also completely ignore the number of yogurts purchased on a particular shopping occasion, avoiding what is in actuality a complicated discrete/continuous choice. Regarding the learning process, our 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. We note that this purchase quantity data may be a potentially interesting source of information on learning for future work.

[31] 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.

[32] This "presample" purchase data is assumed exogenous to our model, and as might be expected are very good predictors of λ_i. In the reduced form models of Ackerberg(1996) other household characteristics such as ages and sexes were not significant. Note that because they do not change over time, coefficients on individual Yoplait 150 observable characteristics X_j such as calories or fat content (in a linear utility, binary choice model) would not be separately identified. WLOG, λ_i represents the sum of the utilities from these characteristics for consumer i.

programming problem and estimation, these econometric unobservables are assumed i.i.d. Type 1 Extreme Value deviates. As in a standard discrete choice model, we cannot identify relative levels or variances of the utility function. The lack of a constant term in the outside alternative utility is our additive normalization; the fixed variance of the ϵ_{it}'s the multiplicative.

As in the theoretical model, δ_{it+1} represents the utility consumer i obtains from the experience characteristics of Yoplait 150 if it is purchased and consumed in period t. Again, this term is generally *not known* before the period t purchase decision. As such (given our linear utility formulation), what is relevant to this decision is the *expectation* of δ_{it+1}, which is determined by the consumer's information, i.e. what they have observed in the past. Recall that a consumer in period t has observed advertising levels of Yoplait 150 in each past period (the sequence $\{a_{i1}, ..., a_{it}\}$), and has observed realizations of experience utility in past periods in which Yoplait 150 was purchased and consumed ($\{\delta_{i1}, ..., \delta_{iK_{it}}\}$). The consumer combines these observations with his priors (as per eqs. (2.7)) to compute his period t posterior means (m_{it}^{δ} and m_{it}^{a}) on the mean experience utility he obtains from consuming Yoplait 150 (δ_i) and how much Yoplait 150 is advertising (a). As the per-period experience utility draws δ_{it+1} are distributed around δ_i, the period t expectation of δ_{it+1} is simply the consumer's period t posterior mean on δ_i, m_{it}^{δ}.[33] Importantly, recall that it is through this posterior on *experience utility* (m_{it}^{δ}) that our informative, signaling effect of advertising influences expected utility. A positive covariance term in initial priors implies that past advertising observations $\{a_{i1}, ..., a_{it}\}$ positively affect m_{it}^{δ} (as well as m_{it}^{a}). In this case the consumer associates or interprets high advertising levels as signaling information on good experience characteristics.

Lastly, note that consumer i's period t posterior mean on Yoplait 150's advertising intensity, m_{it}^{a}, enters directly into the utility derived from consuming the brand. The parameter on this term, θ_3 (and in some cases θ_{3i}) captures our direct, "prestige" effect of advertising on utility. $\theta_{3i} > 0$ indicates that all else equal, the consumer obtains more utility the more he thinks Yoplait is being advertised. In summary, our informative effect of advertising enters expected utility through m_{it}^{δ} ($E[\delta_{it+1}]$), our

[33]This is just the combination of two normal distributions. The consumer knows $\delta_{it+1} \sim N(\delta_i, [\Phi]_{11})$ (but doesn't know δ_i) and "perceives" (has the posterior) that $\delta_i \sim N(m_{it}^{\delta}, [\Sigma]_{11})$. Therefore, he sees δ_{it+1} as coming from the "combination" of the two distributions, i.e. $\delta_{it+1} \sim N(m_{it}^{\delta}, [\Sigma]_{11} + [\Phi]_{11})$.

prestige effect of advertising through m_{it}^a.

From an econometrician's point of view, a consumer's two sources of information (past δ_{it+1}'s and a_{it}'s) differ considerably because while we observe a consumer's advertising exposures, we clearly do not observe his realizations of experience utility. In our model, consumers want to determine how much advertising a brand is engaging in. Thus we define a_{it}, consumer i's observed advertising intensity in a given period t, as the number of advertisements seen by i between the current (t) and previous $(t-1)$ shopping trip *divided* by the amount of television watched during that period[34]. This controls for the fact that different consumers watch different amounts of television. Our econometric assumption on the realized δ_{it+1}'s matches our theoretical one, i.e. that they are assumed distributed normally around δ_i. The variance of this distribution, σ_ν^2, is assumed constant over consumers and time (recall that this variance determines the length of the learning process, e.g. $\sigma_\nu^2 = 0$ implies a one-consumption learning process). 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[35]. 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. While our formulation directly controls for the obvious fact that consumers watch different amounts of television, the data also indirectly suggests that consumers differ in the type of programming they watch (we find statistically significant

[34]As mentioned in the introduction, there are a number of reasons why we define our advertising variables as advertisements per TV hour. First, we think that this definition best corresponds to the specific effects of advertising included in the model. One could argue that both signaling and prestige effects should depend on how much the brand is advertising, not how many advertisements the consumer sees (in contrast, if advertising informed consumers of existence or characteristics, we would expect it to depend more on the absolute number of ads seen by the consumer). Second, this definition may alleviate measurement error resulting from TV meter problems. Third, and perhaps a result of the first or second, Ackerberg (1996) found intensities to do better at explaining the data than absolute number of advertisements. Also note that we are assuming that we observe advertising exposures exactly. We would have preferred to explicitly include advertising measurement error as an unobservable in the model (this method could also be used to deal with the problem of initial advertising in the first 3 months). Unfortunately, as advertising in any given period affects decisions in all future periods, this results in a rather nightmarish integration problem in estimation.

[35]Because the advertising process is inherently binary (either you see or don't see an advertisement for Yoplait 150 in a given minute (or quarter-hour)) this variance computation is a simple function of the probability of seeing an ad in the time segment (assumed constant over time) and the amount of time segments. It should be noted that because a_{it} is essentially the sum of binary random variables , it takes a rather large stretch of the imagination to invoke a CLT and assume its normality in the Bayesian updating rules, particularly in light of the extremely low frequency of advertising exposures (on average one every three or four weeks, although this varies significantly across households). Unfortunately, this is a necessary assumption in developing feasible Bayesian updating rules (Since we need a conjugate *bivariate* distribution with correlation, we are pretty much forced to use normals) . We hope the fact that this is a third-order problem prevents it from affecting the results significantly.

differences in consumers' mean (over time) a_{it}'s , i.e. how many total Yoplait 150 advertisements they saw divided by total hours of television watched over the sample). To accommodate this, we add an additional, consumer level of variance to the advertising exposure process. Specifically, we assume that the a_{it}'s are distributed normally around a consumer-specific advertising intensity a_i (in some sense measuring the type of programming consumer i watches) which in turn are distributed normally around the brand's advertising intensity a[36].

The mean experience utilities of each consumer, δ_i (what the δ_{it+1}'s are distributed around) are themselves econometric unobservables. Because of the linearity of the utility function, we merge any initial expectations the consumer has on δ_i into λ_i, essentially treating consumer i's initial expected value of Yoplait 150's unobservable characteristics as an "observable" characteristic. This normalizes each consumers initial prior on δ_i to zero ($m^{\delta}_{i0} = 0$) and gives δ_i the nice interpretation of being prediction error by the consumer on the utility obtained from consuming Yoplait 150. $\delta_i > 0$ indicates that consumer i is unexpectedly surprised by the quality/taste of Yoplait 150. These δ_i are assumed to be distributed across the population as:

$$\delta_i \sim iid \quad N(\delta, \sigma_i^2) \tag{4.1}$$

independently of λ_i. This independence disallows, for example, the experience characteristics of Yoplait 150 to be (on average) more or less *unexpectedly* preferred by those who like its observable characteristics[37]. Note that we do not fix the mean of δ_i across the population (δ) at 0 - this is a parameter which we estimate and can be interpreted as the overall experience "quality" of Yoplait 150. A δ>0 indicates that on average (across consumers), Yoplait 150 was better than initially expected.

[36]As the a_{it} are observables this does nothing to estimation or the model except that now m^a_{it} 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 posterior on a is just going to follow the posterior of a_i around, so plugging into utility the posterior on a_i rather than a should make virtually no difference (in fact, we believe that the period t posterior on a may be a linear function the posterior on a_i (one can fairly easily show the linear relation holds for 0,1,and ∞ observations of a_{it}) - if this is the case, plugging in the posterior on a makes exactly no difference). On the other hand, the empirical fact that a_i seems to vary across our consumers raises the question of whether a_i may be correlated with consumer specific unobservables in the model. This is discussed below.

[37]We have also chosen not to model δ_i as a function of consumer observables (as is done with λ_i). Both this and the previous assumption (δ_i independent of λ_i) could in theory be relaxed. Identification and computation issues associated with the increase in parameters have prevented this.

The initial prior mean on advertising, m_{i0}^a, and the initial prior variance matrix, Σ_0, are assumed constant across individuals[38] and estimated. Recall that the covariance term in Σ_0 measures our informative effect of advertising. In initial runs we had trouble jointly identifying m_{i0}^a and the 3 individual elements of Σ_0. As a result we have added to the model restrictions derived by assuming that consumers perceive the relationship between experience quality and advertising to be linear, i.e. a brand with experience quality δ advertises according to:

$$a = \beta_0 + \beta_1 \delta \tag{4.2}$$

Then, if $a_i \sim iid \quad N(a, \sigma_a^2)$ independently of the distribution of δ_i around δ[39] and our consumers' initial priors on δ have variance σ_j^2, we obtain the following initial joint prior on δ_i and a_i:

$$\text{Initial Prior:} \qquad \begin{pmatrix} \delta_i \\ a_i \end{pmatrix} \sim N\left(\begin{pmatrix} m_{i0}^\delta \\ m_{i0}^a \end{pmatrix} = \begin{pmatrix} 0 \\ \beta_0 \end{pmatrix}, \quad \Sigma_0 = \begin{bmatrix} \sigma_i^2 + \sigma_j^2 & \beta_1 \sigma_j^2 \\ \beta_1 \sigma_j^2 & \sigma_a^2 + \beta_1^2 \sigma_j^2 \end{bmatrix} \right)$$

Note that there are two components of the consumer's prior variance on δ_i: σ_j^2 arises from uncertainty about Yoplait 150's overall experience quality (δ), σ_i^2 from the household specific deviation in δ_i around δ. As might be expected, and as can be seen in the above formulation, our signaling effect of advertising works only to inform consumers on δ, i.e. the σ_j^2 component of variance. In estimation, we replace σ_a^2 with $\widehat{\sigma_a^2}$, estimated directly from the data, and replace a with $\widehat{a}$, also directly estimated (using (4.2) this defines β_0 given the parameters β_1 and δ[40]). In summary, the linearity assumption

[38] We feel that the most likely potential misspecification of these "constant over consumers" assumptions is that of the initial prior variances on experience utility. One might expect some consumer to have a better idea of the experience characteristics of Yoplait 150 than others. We have thought about modeling heterogeneous prior variances using the "presample" Yogurt/Regular Yoplait data (already used in predicting λ_i), the idea being that the higher these variables are, the smaller is the prior variance. On the other hand, we doubt we could identify these coefficients well and they significantly complicate the dynamic programming problem solution. We don't see an obvious direction of bias in our advertising coefficients resulting from this potential misspecification.

[39] This would not be the case if firms could aim advertisements towards consumers who like the *unobservable* characteristics of the product. This would generate additional covariance in priors. Note that there is not a problem here if firms aim advertisements towards consumers who like the *observable* characteristics of the product (though this brings up additional econometric issues (see footnote below)).

[40] Note that this step (using the observed $\widehat{a}$ to "define" β_0) is the only part of the estimation that requires consumers be *correct* in their beliefs about the relationship between advertising and mean experience utility. We feel (and preliminary results have suggested) that freeing β_0 would not affect our advertising coefficients.

on consumer beliefs reduces the parameters $\delta, m_0^a, \sigma_i^2$, and 3 elements of Σ_0 to the parameters $\delta, \sigma_i^2, \sigma_j^2$, and β_1(Instead of β_1, we actually estimate ρ, the correlation coefficient implied by β_1).

Given that this is a fairly complicated econometric model, 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 the time frame of the sample, in particular after the potential acquisition of information (either through consumption experiences or advertising exposures). 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 (and the resulting accumulation of information) will change a consumer's purchasing patterns. Eventually, everything about the brand is learned and a consumer's purchase patterns will converge to some "post-information" level. σ_λ^2, the variance in the unobserved component of consumers' *known* tastes for Yoplait 150, is identified by unobserved heterogeneity in consumers' "pre-information" (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 the variance of "post-information" heterogeneity[41]. δ, the mean experience utility of Yoplait 150, is assessed by a comparison of the means of these two distributions, i.e. whether "post-information", consumers (on average) purchase Yoplait 150 more or less than "pre-information" (net of experimentation behavior due to dynamic optimization). σ_ν^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 "post-information" level. If, for example, purchase patterns change after initial purchases, but not thereafter, it is indicative that $\sigma_\nu^2 = 0$, i.e. a one-consumption learning process. The advertising-related coefficients, β_1 and θ_3, are identified generally by the effects of advertising exposures on purchasing patterns. Again, the two are separately distinguished by the relative effects of advertising exposures on inexperienced and experienced consumers[42].

[41]More precisely, one wants to compare the variance of post-information heterogeneity to that of pre-information heterogeneity. Under our assumptions (in particular that the random component of λ_i and δ_i are independent and that utility is additive in the two terms), the difference in these variances is σ_i^2.

[42]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).

Moving to estimation, the primary issue is the large amount of 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 our observables $y_i, a_{it}, p_{it}, sc_{it},$ and p_{it}^{oth}[43]. Because of the persistent unobservables and the dependence of purchase probabilities on lagged endogenous variables (through posteriors), we use 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[44]. This results in the following likelihood function (for consumer i) :

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

where c_{it} is the consumer's observed choice in period t, $c(\cdot)$ is the model's predicted choice, $z_{it} =$

[43]We think that 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 amounts of such behavior from occurring. Another potential endogeneity problem arises if firms are able to focus advertising towards consumers who like Yoplait 150 more than our observables predict (i.e. a_{it} is correlated with the random component of λ_i or δ_i) . In Ackerberg (1996), this possibility was tested by including a consumer's mean advertising exposure level ($\approx \frac{1}{T}\sum a_{it}$) as a consumer characteristic. This obtained a positive but insignificant coefficient.

[44]With the exception of the logit errors, the integrals generated by these unobservables are not analytically computable and we rely on either simulation or discrete approximations to evaluate them. As is well known, simulation of these integrals combined with Maximum Likelihood estimation results in inconsistent estimates for a finite number of simulation draws. We do use some crude importance sampling techniques to reduce simulation error and inconsistencies. Note that an alternative to this SML procedure would be to use a simulated GMM framework, using moments in either sequence probabilities (McFadden (1989)) or transition probabilities (Keane (1994)). Because of the large number of potential choice sequences (given about 60 time periods per individual), the feasible version of McFadden's estimator would require numerically simulating the logit errors in addition to the other unobservables. Though in this case one would obtain consistent estimates regardless of the number of simulation draws, this would greatly increase both computational burden (the estimator is not continuous in the parameters) and simulation error, a trade-off we think may not be worth it. Keane's transition probability estimator solves this problem, but like SML, is not consistent for a fixed number of simulation draws. While Keane (1994) shows that small sample bias is small (relative to SML) for short panels ($T = 8$), it is not clear what happens in very long panels when the denominators in his estimator will likely contain much more simulation error.

$(p_{it}, p_{it}^{oth}, sc_{it})$, T_i is the total number of shopping trips of consumer i, and superscripts indicate histories of a variable through that point (e.g. $a_{it}^t = \{a_{i1}, ..., a_{it}\}$). 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, conditioned in particular on λ_i, past choices c_{it}, and past realized δ_{it+1}'s[45]. The predicted choice function $c(\cdot)$ is defined by either myopic utility maximization (i.e. $I(EU_{i1t} > EU_{i2t})$) or the optimal policy function generated by the dynamic programming problem. Note that this choice function is a function of current posterior means $m_{it}(\cdot)$, which in turn are a function of past realizations of experience utility and past advertising intensities. 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 (see Rust (1987)).

4.2. Dynamic Programming Solution

In estimating the full dynamic model, we must solve the consumer's dynamic programming problem to obtain the predicted choice function $c(\cdot)$. 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. Our utility specification and the assumed learning process generate the Bellman's equation:

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

where the state space $s_{it} = (\lambda_i, m_{it}, \Sigma_{it}, p_{it}, sc_{it}, p_{it}^{oth}, \epsilon_{iot}, \epsilon_{i1t})$. Although this state space appears to be quite large, there are a number of simplifications and assumptions that we use to significantly reduce the dimensionality of the problem and allow for relatively quick numeric solution. First, conditional on 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, our assumption that Σ_0 and Φ are constant across consumers[46] allows us to replace Σ_{it} in the state space by K_{it} and TV_{it}. Second,

[45] Note that formally we are integrating over the T_i *potential* δ_{it+1}'s on *every* shopping trip of consumer i (not just trips in which a Yoplait 150 purchase was made). In practice, since the δ_{it+1}'s on non-purchase trips are not observed by the consumer and don't affect anything, these integrals disappear, leaving just integrals over the δ_{it+1}'s associated with actual purchases.

[46] For dynamic programming purposes (to get Φ constant), we assume that all consumers *anticipate* watching the same amount of television between the current and next shopping (fixing this amount at the sample mean). Note that this assumption *only* affects perceptions of the future - in the actual Bayesian updating formulas used in likelihood evaluation,

because δ_{it+1} enters the utility function linearly, λ_i can be merged into the learning process (so our 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 (Pakes (1986), Rust (1987), Keane and Wolpin (1994)). This removes the i.i.d. ϵ's from the effective state space, and as we assume that consumers perceive p_{it}, sc_{it},and p_{it}^{oth} to be non-serially correlated[47], 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 ϵ's in the Bellman equation can be computed analytically (Rust(1987)). This, along with the assumption of discrete perceived distributions of future p_{it}, sc_{it},and p_{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 in order to apply the Method of Successive Approximations and numerically solve the above dynamic programming problem. Note that we do not discretize the entire dynamic problem, but choose points at which we will solve for the (approximate) value function. In the following estimation results we have discretized the state space into 10 to 20 points in each dimension. Because the numerical integration mentioned above is only two dimensional we chose to use quadrature type methods rather than Monte-Carlo. Since the quadrature points generate future states that are not our discretized ones, we use linear interpolation to evaluate the value function at these states. Unfortunately, the computation burden of even this simplified problem is large enough to prevent a detailed exploration of either alternative numeric methods or the

Φ depends on the amount of television watched between $t-1$ and t. A similar assumption involves the discount factor. Although time between shopping trips varies significantly over consumers and time, we assume that consumers use the same per-shopping trip discount factor to weight the future. Again, this only affects perceptions of the future. Relaxing either of these assumptions would at the very least add an extra "state" variable to the model (i.e. the time until your next shopping trip, the amount of TV you will watch between now and then)

[47]We assume a 4-point (estimated from the data) perceived distribution of future p_{it}, and 1-point expectations of sc_{it} and p_{it}^{oth}. We note that the assumption that the p_{it} are iid is easily rejected by the data, and is only adopted for computational reasons. We believe that this effect of this on estimation should be small, as this distribution only enters into expectations of the future in the dynamic program (and it doesn't enter the myopic estimation results at all). See Ackerberg (1996b) for a more thorough development of the concepts of this section in the context of our model.

sensitivity of the parameter estimates to the current discretization process[48].

4.3. Results

Table 4.1 presents maximum likelihood estimates of the above model. In initial runs we had trouble obtaining reasonable estimates of σ_ν^2, the per-period variance in experience utility. In particular, our estimates were unreasonably high, indicating that consumers were learning (through consumption) about the unobservable characteristics of Yoplait 150 very slowly. As this parameter is identified by changes in consumer purchasing patterns after consumption experiences, it is likely that it picks up variables unobserved by us that cause such changes (e.g. learning about other brands or products)[49]. As a result, the majority of our estimates assume a one-period learning process ($\sigma_\nu^2 = 0$) where one consumption experience with Yoplait 150 informs the consumer exactly the value of its experience characteristics to him. This assumption has the added benefit of greatly reducing both the computational burden of the dynamic programming problem and the dimensionality of the integrals in likelihood evaluation. We have capitalized on this computational reduction to increase the precision and accuracy of our discretization and integral evaluation over what would have otherwise been possible[50].

The first two columns of Table 4.1 contain results under the myopic assumption on behavior, with and without a time trend[51] on the outside alternative. The last three columns are results from

[48]For example, one could move to higher order interpolations, as perhaps this can obtain more accuracy with a lesser number of discretized points (Rust(1995), Johnson et. al.(1993)). Rust (1997) also suggests randomly picking state space discretization points in order to break the curse of dimensionality. One could also move to pure Monte-Carlo simulation or importance sampled Monte-Carlo for the expectation. We felt that conditional on having only ˜50 pts to work with, discretized normals would be better. Informal experimentation with our number and choice of discretized points results in slightly different estimates, but found no evidence to alter the conclusions of this study.

[49]We leave the investigation of potential misspecification due to not allowing such effects to future work.

[50]With this assumption, the state variable "number of previous Yoplait 150 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 3-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 26ˆ2 draws), one function evaluation with a discount rate of .98 on an UltraSparc in highly optimized C code takes about 7 minutes ($\sim$5.5 min. for the contraction mapping to converge ($\sim$1 sec. per contraction iteration), 1.5 min. for likelihood evaluation). Even with good starting parameters, the entire maximization procedure can take up to a month.

[51]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. This would not be the case

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 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 .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. In the dynamic models including a time trend, reasonable values of the discount factor result in worse likelihoods than the corresponding myopic model. This is likely the result of the positive time trend on the outside good generating behavior similar to the experimental behavior generated by a higher discount factor. We therefore fix the discount factor at its point estimate from the dynamic model without the time trend.

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.

if we had, for example, a quadratic time trend.

The estimated coefficients pertaining to advertising strongly support the hypothesis that advertising is affecting consumers *only* through the informational structure of the model. Our "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. On the other hand, the estimates of ρ, the correlation coefficient corresponding to Σ_0, are all significantly positive, suggesting that advertising is providing consumers with information. The differing levels of the estimates of ρ across the columns do not reflect differences in the magnitude of this informative effect of advertising. ρ is not as orthogonal a measure of the informative advertising elasticity of demand as we would like. This elasticity depends heavily on the relation of ρ to the overall magnitude of Σ_0, which is not identified particularly well and varies over the sets of estimates. The parameters at the bottom of the table provide a better indicator of this informative effect of advertising. 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 (To focus on the informative effect we do not double the advertising level that enters the prestige effect). 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 .15[52]. These are very similar to those found in Ackerberg (1996) and along with the estimated price elasticity imply an advertising to sales ratio of 4.5% in a static, single-product firm, advertising and price-setting model[53].

[52] That is, doubling Yoplait's advertising level over the entire time frame of the data results in a 15% increase in sales. We don't approximate standard errors here because the derivatives needed for the delta method are extremely hard to simulate accurately for the entire sample (in contrast to those for a representative consumer above). We expect that these overall elasticities (as well as overall *informative* elasticities) would also be significant.

[53] One can obtain this result by differentiating the profit function $\Pi(p,a) = q(p,a)(p - mc) - c_a a$ where $q(p,a)$ is demand as a function of price and advertising, mc is the marginal cost of the product, and c_a is the cost per "unit" of advertising intensity. This simple model combined with our estimates also predicts a price-cost markup of 30%. Though these seem to be reasonable results (According to Advertising Age, in 1988 total Yoplait advertising expenditures were about 7% of total sales) and provide ballpark figures, this simple firm side model is obviously deficient. First, with our demand environment, firm behavior should be dynamic in nature. Second, this is not a single-product firm. Third, as we don't model consumer's purchase amounts, these elasticities are not necessarily the elasticity of total quantity (units) purchased with respect to price (or advertising).

Table 4.2 examines three perturbations of the myopic model (these perturbations make the dynamic model infeasible to estimate). In the first column we allow for a more flexible specification of the persuasive effect of advertising by including dummies for m_{it}^{a} lying in different regions. The estimates are generally imprecise and don't follow any general trend. We have also tried this simultaneously estimating regions and using non-linear functional forms, but haven't found much. The estimates in the second column include a random coefficient on persuasive advertising, addressing the possibility that there may be heterogeneity in prestige or image effects across the population. This random coefficient is constant for a given consumer through the time frame of the model and assumed independent of the observables and other unobservables[54]. Although the estimated standard deviation of the random coefficient is economically large it is insignificant even when simulation error is neglected[55]. Note that the mean prestige effect stays at essentially zero. We hesitate to make any strong conclusions about these results because the random coefficient may be picking up measurement error in our advertising variable. The third column relaxes our assumption of a one-period learning process, fixing σ_{ν}^{2} arbitrarily at 1 (This results in a path of posterior experience quality variances (in # of prior purchases) of 2.882, .742, .426. .299, .230, .187,....). Note that in this model our informative effect of advertising can influence a consumer's behavior after their first consumption experience (though its effect diminishes in number of prior consumption experiences). Although the parameters of the learning process change significantly in this set of estimates, we still obtain an insignificant prestige effect and a positive, significant informative effect.

Before proceeding, we reiterate some important points that arise from our inability to include or empirically distinguish more than one informative effect of advertising in our model. First, we feel alternative models with different formulations of the informative effect of advertising discussed in the introduction would produce similar empirical results, i.e. a significant informative effect, insignificant

[54]With more data, it might be interesting to allow this random coefficient to depend on consumer characteristics. Note that in the dynamic programming problem, this random coefficient would be an additional "effective state variable" (i.e. the value and choice functions would depend on it).

[55]For a consumer whose prestige random effect is 1 standard deviation away from the mean, a doubling of advertising at the mean has the same effect on purchase probability as a 5 cent price decrease. Because there are 3 unobservables to integrate out over in this model, we switch from discretizing the integrals to using Monte-Carlo with crude importance sampling. We also use this technique in the next set estimates, where there are $(2+k_{iT})$ dimensional integrals to evaluate. As a result of this, the likelihood values in the last two columns are not directly comparable with the previous likelihoods.

prestige effect. Underlying this point is a strong belief that the differential effect advertising on inexperienced and experienced consumers that separately identifies the informative and prestige effects of advertising in the current model would provide similar distinction of informative and prestige effects under alternative informative effect formulations. As a result, we interpret our results as strong empirical support for: (1) the conclusion that there were no Becker-like prestige effects generated by these advertisements, and (2) the hypothesis that 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[56].

There is a relatively clear cost-benefit trade off with our signaling effect of advertising: costs the cost of resources devoted to advertising, benefits the information conveyed to consumers by the advertising (we ignore potential benefits (or costs) of advertising outside this market, in particular its subsidization of media.) On the other hand, for possible prestige effects, there is a serious question as to how to measure potential benefits of advertising. Because utility is a latent variable, a positive prestige effect only indicates that advertising increases the utility of consuming Yoplait 150 *relative* to other yogurts. We cannot distinguish whether it adds to the utility obtained from consuming Yoplait 150 or subtracts (i.e. dis-prestige) from the utility derived from consuming the outside alternative[57]. As this distinction would clearly be relevant to any welfare analysis, our conclusion that there is no

[56]Not only is it dependent on this strong assumption, but also on numerous and likely simplistic firm behavioral assumptions and equilibrium extrapolations that will follow. As a result we certainly do not take this to be anywhere near an "end all" conclusion on the welfare effects of advertising in this market. On the other hand, we do feel is that it is an interesting exploration of the structural estimates and illuminate interesting (and typically ignored) possibilities concerning the welfare effects of informative advertising. Essentially, we push our structural model to the "bitter end". What is additionally of interest is just how far this "bitter end" is.

[57]Nor anything in between, nor for that matter if it adds to or subtracts from *both* utilities in different amounts. This result is in theory true in any discrete choice model, for example, one might attempt a similar argument with price or product characteristics. In that case, however, there is no reason why the price or characteristics of Yoplait 150 should have any effect on the utility obtained from not consuming Yoplait 150. With the intangibility of prestige effects, this line of argument is much more believable.

prestige effect of advertising in this market is quite facilitating here[58].

Our first step is to combine the final dynamic estimates of the demand for Yoplait 150 (the last column of Table 4.1) with profit maximizing first order conditions of Yoplait to "back-out" the production cost of Yoplait 150 and costs of advertising. Unfortunately, the first-order conditions which we would like to have, those arising from Yoplait's dynamic price and advertising setting problem, are not feasibly obtainable within the context of the fairly complicated consumer side model presented here. Even ignoring other products, the firm's state space for such a dynamic model would need to contain the joint distribution of consumer tastes and experience. This is far too complicated for the present work and we proceed using a major simplification of producer behavior: one in which the firm sets *one* price and *one* advertising level in order to maximize profits in some "introductory period"[59]. Another problem is that we do not observe a single price, but rather a distribution of prices. Again in order to simplify things, we assume that the firm chooses a mean price with the price distribution around that mean price fixed (as per our sample - perhaps due to retailer behavior). Importantly, note that this mean price is the average price observed by a consumer on any given shopping trip, *not* the average price paid for Yoplait 150 (which is presumably lower). While we concentrate on reporting how $\overline{p}$ changes under alternative policies, our welfare calculations depend on optimal consumer response to these prices and thus depend more directly on the prices actually paid.

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

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

[58] Were this not the case we might do something like the following: Suppose we estimated a positive prestige effect θ of Yoplait 150 advertising. It might be reasonable to assume that this positive *relative* effect could not decrease the absolute utility from consuming Yoplait 150 and could not increase the absolute utility from consuming the outside alternative. This puts bounds on the effects on un-normalized utility at $(0,\theta)$ on Yoplait 150 (respectively $(-\theta,0)$ on the outside alternative). One could then do a welfare analysis on different points in this range (in particular the best and worst case scenarios for advertising). This is somewhat reminiscent of the procedure in Dixit and Norman's (1987) theoretical work on advertising, but actually quite different. Their range corresponds to the degree to which consumer's are rational in their decision-making (In our model the bottom of this range would entail a consumer making decisions based on U's but receiving utility $U - \theta m_{it}^a$). They do *not* address the possibility that a brand's advertising might affect the utility derived from not consuming a product.

[59] We take this introductory period to be the length of our data. One possible justification for this could be that product qualities are somehow revealed to consumers after the introductory period, essentially removing any dynamic effects of current price and advertising setting on profits after this period.

where $q(\overline{p}, a)$ and $TR(\overline{p}, a)$ 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. $TR(\overline{p}, a)$ and $q(\overline{p}, a)$ are quantities which we can simulate using our estimated structural model. This involves drawing consumers from the estimated distribution of consumer heterogeneity, simulating prices and advertising exposures, and computing information paths and optimal purchase decisions through the time frame of the model. Note that TR cannot be decomposed (into $p \cdot q$) because of the distribution of prices. Differentiating with respect to both $\overline{p}$ and a and manipulating the two F.O.C.'s gives:

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

Using our estimates we simulate these four derivatives and solve these equations, obtaining a marginal cost of Yoplait 150 of \$0.422 ($\overline{p} =$ \$0.653) and a cost per unit of advertising intensity of \$3732[60,61].

With these costs in hand, we can analyze the welfare effects of a ban on advertising[62]. We 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 be

[60]This advertising cost is not normalized (i.e. not per consumer) and thus scales with our simulated dataset of 1.2 million consumers over 20 periods. More generally, because of our assumptions of 1) a constant marginal cost of production and 2) that the firm faces a constant price per unit of advertising intensity ($TC(a) = c_a a$), changing the size of the simulated dataset would change nothing (except for the scale) of our calculations (if we double population, then q, TR, and their derivatives all double, mc remains the same, and thus c_a doubles). Note that one could argue that due to supply side conditions (i.e. the "zero marginal cost" nature of the literal production of advertising exposures), the advertising market equilibrium c_a *per person* could actually change as population changes (e.g. because of its large population, advertising in the US may cost less per exposure than in a smaller country). Even if this is the case, the (presumed optimal) advertising we observe is Yoplait's U.S. national advertising. Thus, the c_a *per person* we actually back out using our estimates is the correct cost for U.S. national advertising, and our welfare calculations are correct (scaled appropriately) for a country the size of the U.S. (though perhaps not for a smaller country where c_a *per person* could be different). Obviously these results also rely on the demand curves of consumers in Sioux Falls and Springfield being representative of the U.S. population in general.

[61]In contrast to the simulations in the previous section to obtain price and advertising elasticities, here we redraw advertising exposures according to the estimated advertising process. This smooths out the advertising exposures and surely increases our projected value of the advertising information (The problem in the actual data is the fact that there is no advertising for the first three months drives consumer's priors incorrectly downward, in some cases driving the value of the information negative).

[62]We focus on a total ban (rather than intermediate policies such as taxes or subsidies) because it is the most straightforward analysis. With the structure on the consumer side, we can easily adjust consumer beliefs for any alternative signaling equilibria (i.e. any combination of β_0 and β_1). What we don't have (without a complete structural model of firm behavior) is the map from alternative tax policy into a new signaling equilibrium (except for the case of a ban on advertising where clearly β_0 and β_1 must become zero).

telling consumers anything about Yoplait 150's experience quality. Note that 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 $\beta_0 = \beta_1 = 0$ we obtain a new demand system ($TR(\cdot)$ and $q(\cdot)$ functions), which is used to numerically find the new $\overline{p}$ which solves the above first order conditions (unfortunately, we cannot optimally adjust competitors' prices because we have no model of demand for 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 ($\delta = .899 > 0$). The advertising ban prevents Yoplait from signaling this through advertising and reduces their ability to price this quality.

The first two columns of Table 5.1 exhibit some welfare measures of the two equilibria. Both 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[63]. 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 additionally considers the case when Yoplait is not allowed to optimally 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. In sum, this analysis suggests that the costs of advertising Yoplait 150 far exceeded its signaling information 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 we have assumed and estimated is really an equilibrium over many possible new products with many different experience qualities δ[64]. 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

[63]The increase in profits under the ban suggests that in equilibrium ability to advertise might dissuade entry. This goes against the standard argument (see, e.g., Tirole (1988) pg. 288) that informative effects of advertising should promote entry by reducing natural information barriers.

[64]Again, we acknowledge the significant degree of extrapolation going on in this exercise and that our results depend crucially on this (as well as our other very stringent assumptions on firm behavior). In essence, we have estimated an entire equilibrium over potential entrants from data on just one entrant. On the other hand, we find it intriguing how (under our assumptions) this entire equilibrium (or a least the local slope of it) can be identified by the variance in advertising exposures we observe *over consumers* .

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 integrate out welfare benefits over the distribution of *all* possible product introductions (i.e. all possible experience qualities)[65]. Note that assuming consumers are "correct" (in a probabilistic sense) in their priors, the distribution of possible experience qualities that we want to integrate over is actually part of our consumer model - it is the consumers' initial prior on δ (which we have assumed normal, normalized its mean to 0, and estimated the standard deviation $\sigma_j = .593$).

At this point, one might think we are stuck, as we have no direct data on these other "products" with different δ's. On the other hand, given $(\overline{p}, a)$, we can compute (simulate) demand for alternate experience qualities $(TR(\overline{p}, a, \delta), q(\overline{p}, a, \delta))$[66]. In addition, although we do not know marginal costs and optimal mean prices for alternative experience qualities ($mc(\delta)$ and $\overline{p}(\delta)$), our 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 $\overline{p}$), we can numerically solve out these unknowns for each δ. Knowing $mc(\delta)$, we can also invert out optimal mean prices for each experience quality under an advertising ban and the resulting adjusted demand system,. Figure 5.1 plots the results: $\overline{p}(\delta)$, $mc(\delta)$, and $\overline{p}^{noad}(\delta)$ - all increase in experience quality[67]. 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 5.2 plots CV under the two regimes, as well as for the case where the firms do not change

[65] One can interpret this either of two ways: 1) that this is the a-priori expectation (over the realization of the exogenous, random experience quality δ of Yoplait 150) of the benefits of Yoplait 150's introduction, or (2) that new products with random, exogenous experience qualities are entering the market consecutively and this is an integral of benefits over time/products.

[66] Note that in the interpretation of consecutive product introductions, demand also depends on potentially different observable characteristics of these products. In this case we assume that the other products have the same observable characteristics (encompassed in the λ_i's) as Yoplait 150. We also need to assume that the distribution of prices around $\bar{p}$ is fixed.

[67] The fact that prices vary over experience quality raises the question of why consumers can't simply infer quality from price. One possibility is that there is too much variation in the price distribution (although there is variance in advertising also). Another is that, as in Milgrom and Roberts, firms in equilibrium need to set both prices and advertising levels appropriately for a credible signal. A more logistical problem is that our linear advertising equilibrium equation indicates low δ_j's (those more than 1.54 standard deviations below the mean) should advertise negative amounts. In simulating welfare for these qualities, we assume that no money is spent on advertising but that consumers get the correct signals anyway, again likely biasing our results slightly in favor of advertising.

price under the ban (again to separate out information effects from price effects). As expected, the further experience quality is away from their prior mean the more consumers benefit from the advertising information (Figure 5.2b). In fact, for experience qualities close to 0, the effect of advertising information on compensating variation is negative because the variance in the advertising signal moves consumers away from initially close-to-correct priors[68]. 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 δ =1.4 they may not).

Figures 5.3 - 5.6 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[69]. 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[70]. More appealing is the fact that the advertising ban slightly accentuates the slope of the profit curve. This suggests that the ability to advertise increases incentives (or at least decreases disincentives!) to invest in experience quality in a model where such choices were endogenized. Figure 5.7 plots total surplus under the two regimes; only in the very

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

[69]This increase in profits under the ban exemplifies the dependence of these results on the information being signaling information. Suppose that instead, Yoplait 150's advertisements solely informed consumers of its existence on the market. Using the same F.O.C.'s as above, we would never obtain the increase in profits resulting from an advertising ban that we find above. The reason is that such a ban wouldn't affect the "zero advertising" demand curve (as long as consumers don't, for example, increase search) so Yoplait 150's optimal (though now unacheivable) advertising level would remain the same with the ban (i.e. it would still *want* to advertise a positive amount). Therefore, profits must be lower under the ban. This contrasts with the situation above in which the ban on advertising changes consumers' equilibrium beliefs and thus changes the demand curve the firm is facing.

[70]Generating this result is the fact that lower experience quality products' lower prices generate many more first-time purchases and many more idiosyncratic taste draws from a fairly high variance distribution. We suspect that this slope might disappear or change sign if we allowed firms some flexibility in changing prices over time. This would allow higher quality firms to start with a lower prices to stimulate experimentation, then charge a higher price to reap profits. Of course it would also allow low quality firms more flexibility also, but we anticipate the increase in profits would be proportionately less. We are currently attempting to invert out optimal choices when we allow firms to choose two prices (this involves simultaneously solving 3 F.O.C.'s)

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, we feel our structural estimation results are strong and interesting. We present a model in which we explicitly include two effects of advertising: an informative effect which enters the information structure of our dynamic consumer learning model, and a prestige or image effect entering directly into consumers' utility functions. Structural estimation of this model finds a large, significant, and robust informative effect of advertising and an insignificant prestige effect, suggesting that these Yoplait 150 television advertisements affected consumers primarily through the provision of information, not through prestige or image effects. These results support the conclusions of Ackerberg (1996) and strengthen them by explicitly allowing and controlling for experience characteristics and consumer learning. We feel that together the two approaches provide a broad framework within which one can analyze effects of advertising for other products, given the appropriate data. Of particular interest might be comparing estimates across different types of products, seeing if one can find the existence of prestige or image effects.

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

These deficiencies point to further research. One direction is moving to consumer levels models incorporating multiple informative effects of advertising. As suggested in Ackerberg (1996), different informative effects can potentially be distinguished with the proper data[71]. Perhaps more challenging is developing realistic empirical models of firm behavior in markets with imperfect information and advertising. Such models would need to be dynamic, as decisions such as price have dynamic effects though their effects on consumer information. We also would want such models to endogenize entry and innovation. Unfortunately, it is likely not feasible to embed a consumer demand model as rich as the above into a dynamic model of firm behavior. Therefore, the challenge is to develop a demand side rich enough to accommodate such decisions and effects but parsimonious enough to be able to solve and estimate.

[71] Identification might come from two sources - first, as mentioned earlier, one could exploit the difference between advertising exposures and advertising intensities, hypothesizing that more direct forms of information depend on absolute number of exposures, indirect forms of information such as signaling depending on intensities. A second source could be heterogeneity in response to advertising. Explicit information on characteristics might result in heterogenous responses, depending on whether the consumer is favorable towards the explicit information being conveyed.

Table 3.1: Summary Statistics

Variable	Market 1	Market 2
Households	950	825
Total Shopping Trips	67051	54308
Average Shopping Trips per Household	70.58 (33.39)	65.82 (31.82)
Average Price of Yoplait 150 (Cents)	0.645 (0.060)	0.663 (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	12918	12563
Commercial Exposures per Household	13.60 (10.81)	15.22 (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

Table 4.1: 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
Note: Below parameters not estimated, standard errors obtained by delta method					
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)

Table 4.2: 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 (0.33081)	-5.62280 (0.32840)	-5.72480 (0.33384)
θ_2 - Store Coupon	3.12930 (0.81223)	3.06820 (0.82226)	3.11890 (0.81513)
θ_3 - Prestige Advertising		-0.00803 (0.05651)	-0.03285 (0.04685)
θ_4 - Competitor's Price	-0.69068 (0.22560)	-0.71465 (0.22015)	-0.72726 (0.22448)
θ_5 - Time Trend on Outside Alternative	1.23630 (0.16740)	1.07650 (0.17547)	1.28130 (0.19354)
σ_i - Variance of δ_i around δ	1.76940 (0.14222)	1.48520 (0.14672)	2.30260 (2.59952)
σ_j - Consumer's percieved variance of δ	1.45280 (0.14235)	1.70980 (1.38625)	4.90710 (1.64029)
ρ - Correlation Coefficient of Σ_0 - Informative Advertising	0.62582 (0.04930)	0.74675 (0.27701)	0.79062 (0.34281)
δ - Mean Experience Quality of Yoplait 150	-0.75524 (0.09484)	-0.84418 (0.76762)	1.72740 (0.15174)
Discount Factor	0	0	0
λ_i - Constant	-3.43571 (0.44095)	-3.22620 (0.85789)	-3.01320 (1.30272)
σ_λ	1.76680 (0.15519)	1.74530 (0.11035)	1.73290 (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 Coefficient on θ_3		0.11258 (0.07562)	
Log Likelihood	-3940.0390	-3939.6724	-3921.7334
Note: Below parameters not estimated, standard errors obtained by delta method			
Informative Advertising Effect - 1 Week	0.34616 (0.03477)	0.38097 (0.17062)	0.53471 (0.23470)
Informative Advertising Effect - 5 Weeks	0.38876 (0.04632)	0.41891 (0.18613)	0.58903 (0.21620)
Informative Advertising Effect - 20 Weeks	0.41421 (0.05497)	0.44082 (0.20523)	0.62047 (0.23471)
$[\Sigma]_{11}^{1/2} = (\sigma_i^2 + \sigma_j^2)^{1/2}$	2.28941 (0.14028)	2.26478 (1.04530)	2.88182 (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 4.1. Not all λ_i terms shown.

Table 5.1: 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	236371.00	225498.00	218377.00
Total Revenue	148360.93	140129.22	137047.13
Production Costs	99762.744	95173.686	92168.197
Advertising Costs	9329.996	0	0
Total Costs	109092.74	95173.686	92168.197
Profits	39268.186	44955.529	44878.933
Compensating Variation	45541.838	46554.529	44962.923
Total Welfare	84810.024	91510.058	89841.856

Figure 5.1

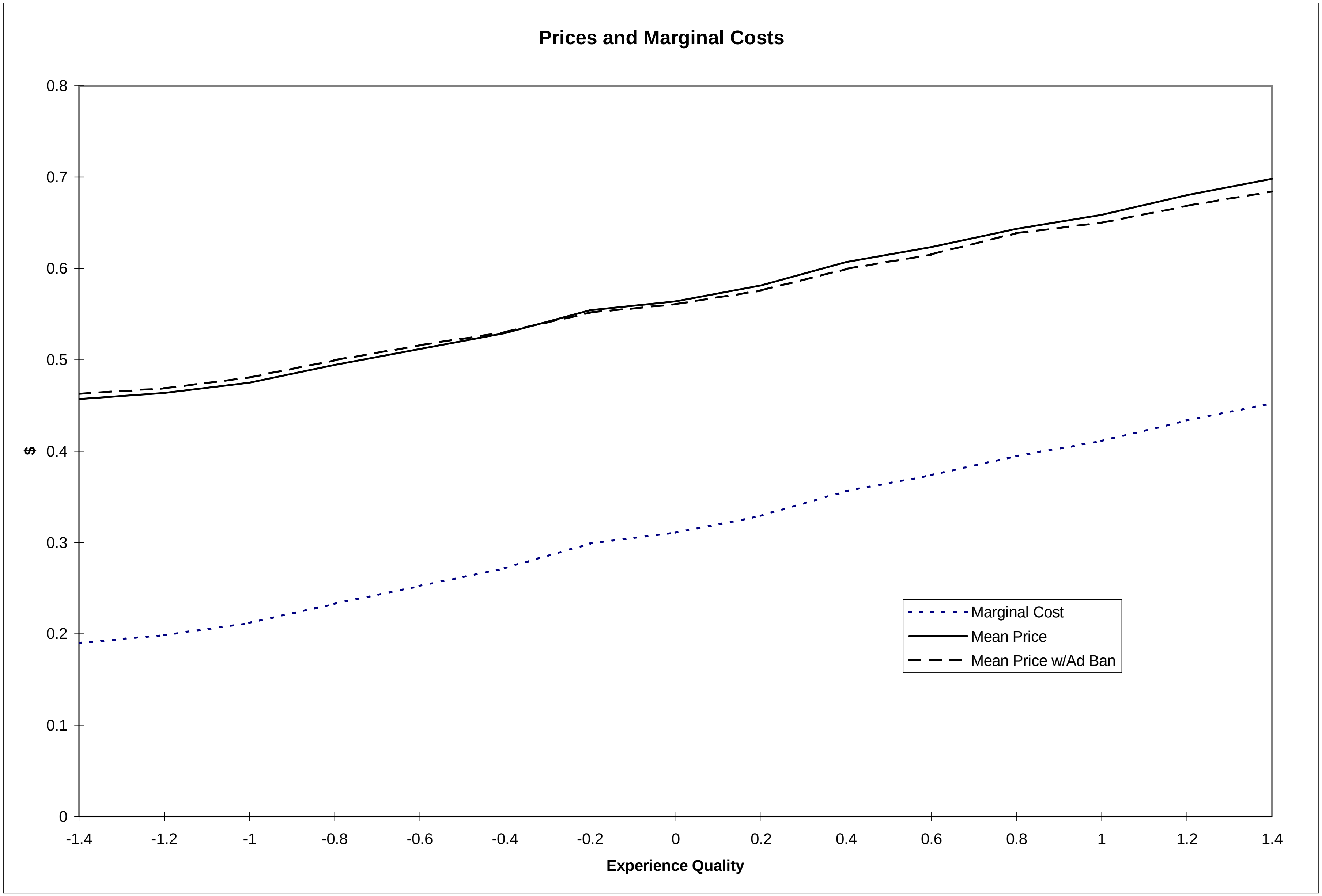

Prices and Marginal Costs
$
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-1.4
-1.2
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
1.2
1.4
Experience Quality
Marginal Cost
Mean Price
Mean Price w/Ad Ban

Figure 5.2

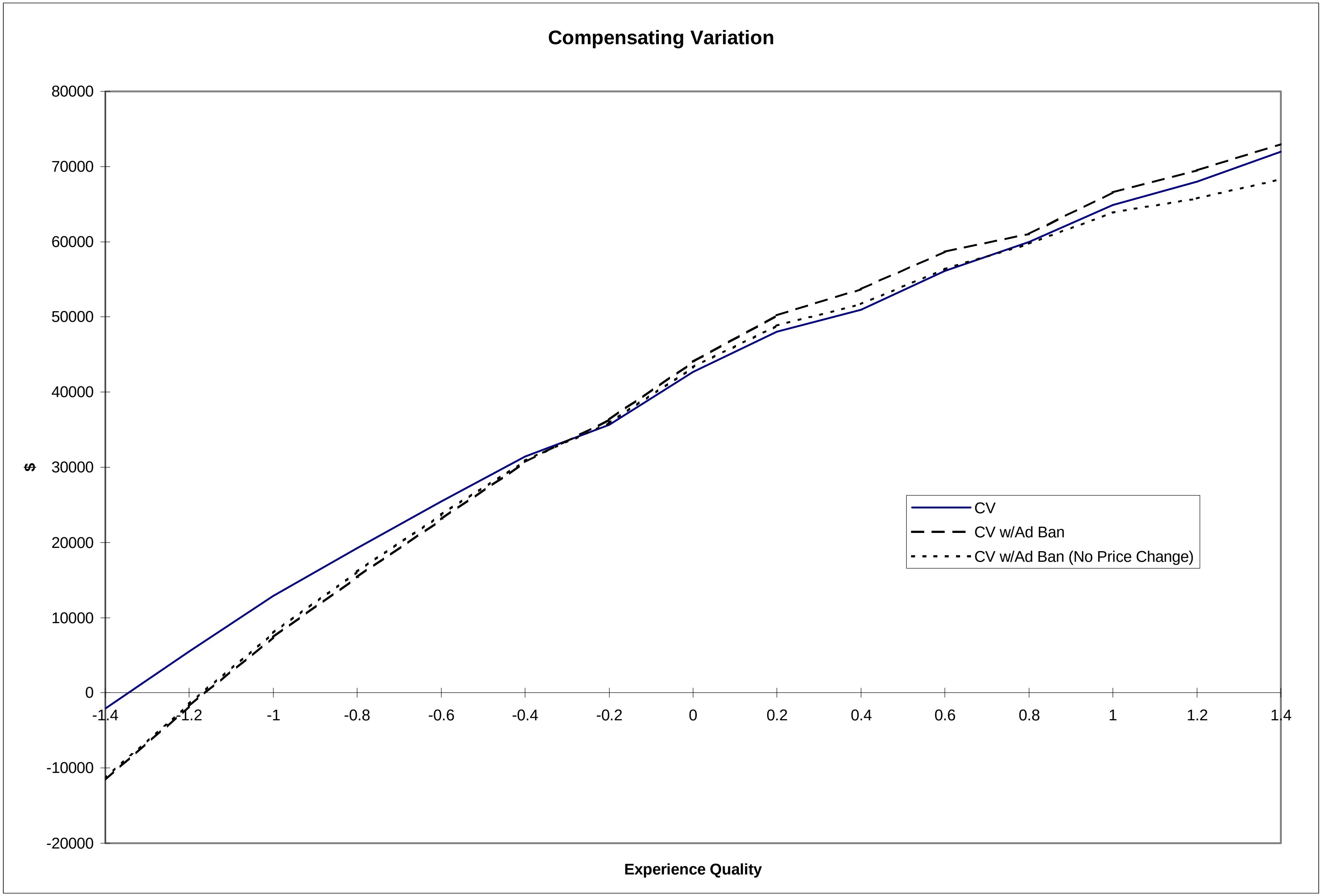
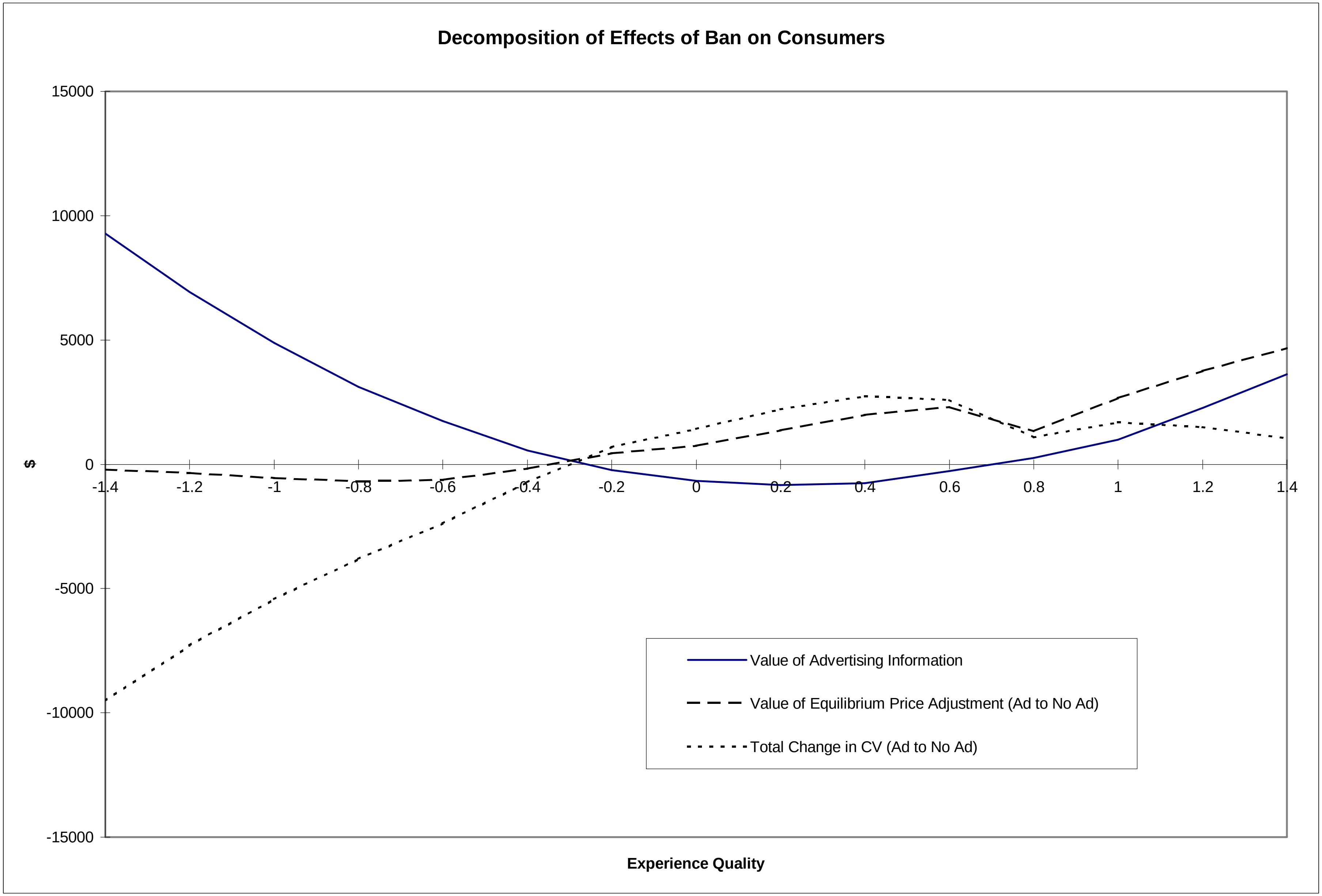
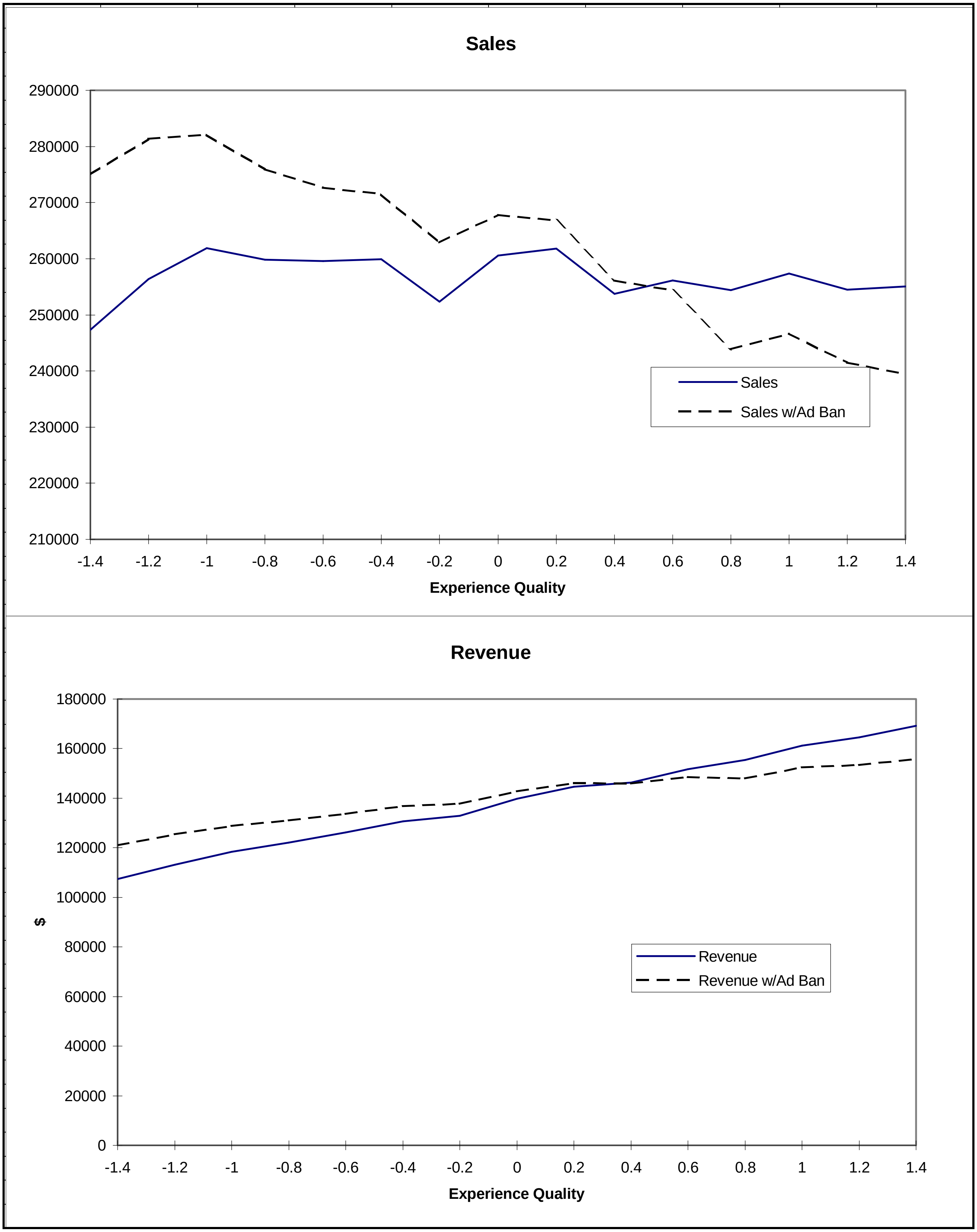

Compensating Variation
80000
70000
60000
50000
40000
30000
20000
10000
0
-10000
-20000
$
-1.4
-1.2
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
1.2
1.4
Experience Quality
CV
CV w/Ad Ban
CV w/Ad Ban (No Price Change)

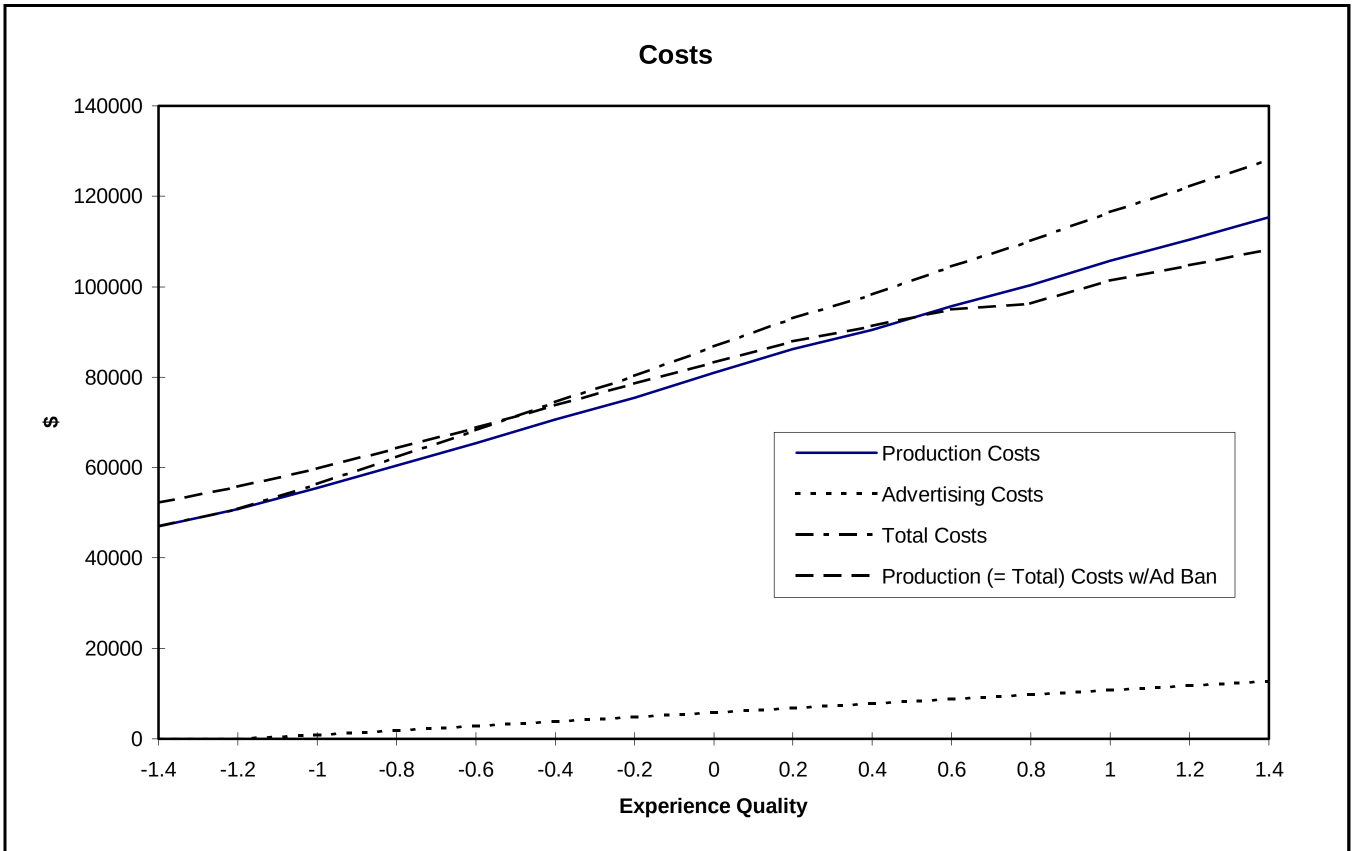
Figure 5.2b


Decomposition of Effects of Ban on Consumers
15000
10000
5000
0
-5000
-10000
-15000
$
-1.4
-1.2
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
1.2
1.4
Experience Quality
Value of Advertising Information
Value of Equilibrium Price Adjustment (Ad to No Ad)
Total Change in CV (Ad to No Ad)

Figures 5.3 and 5.4



Figures 5.5 and 5.6



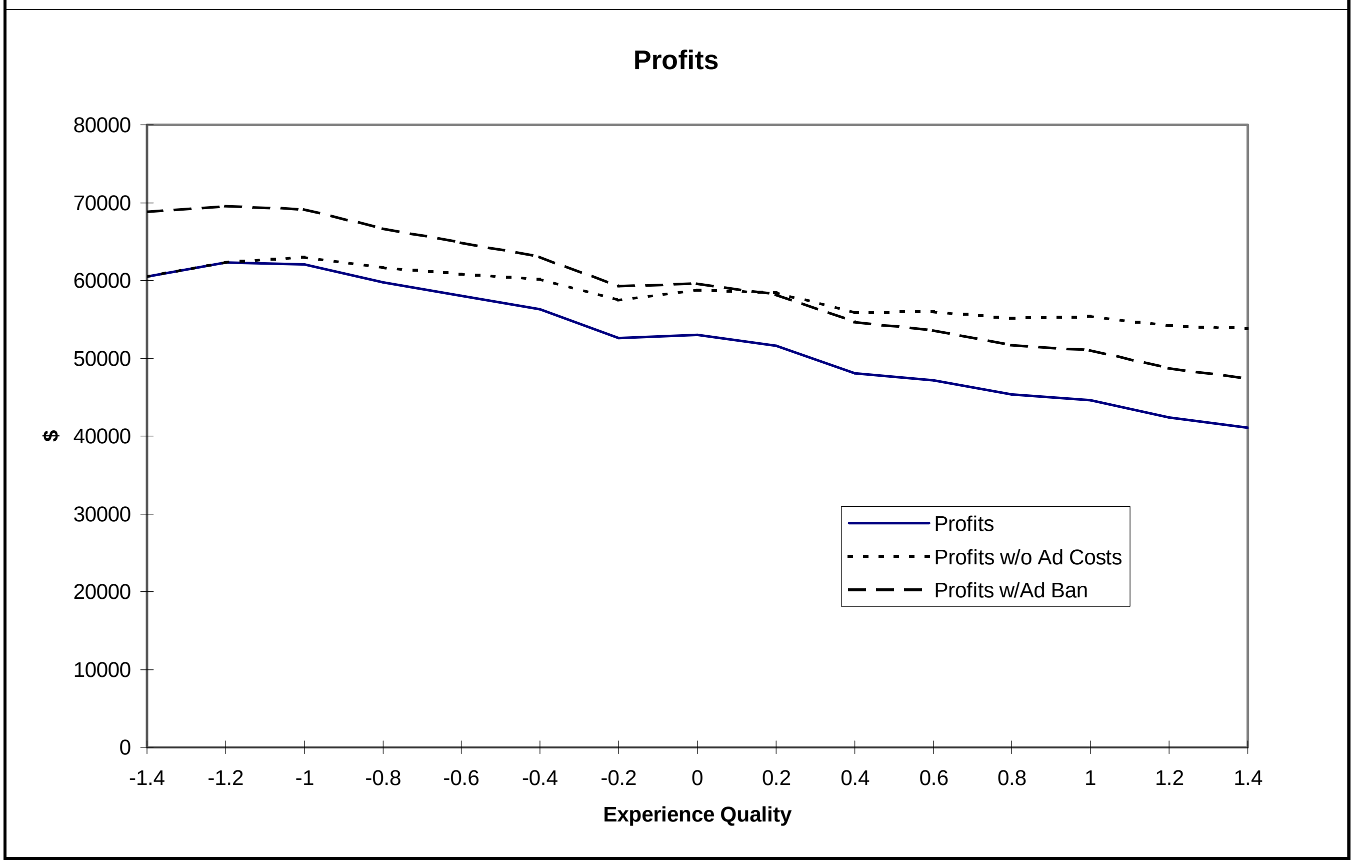
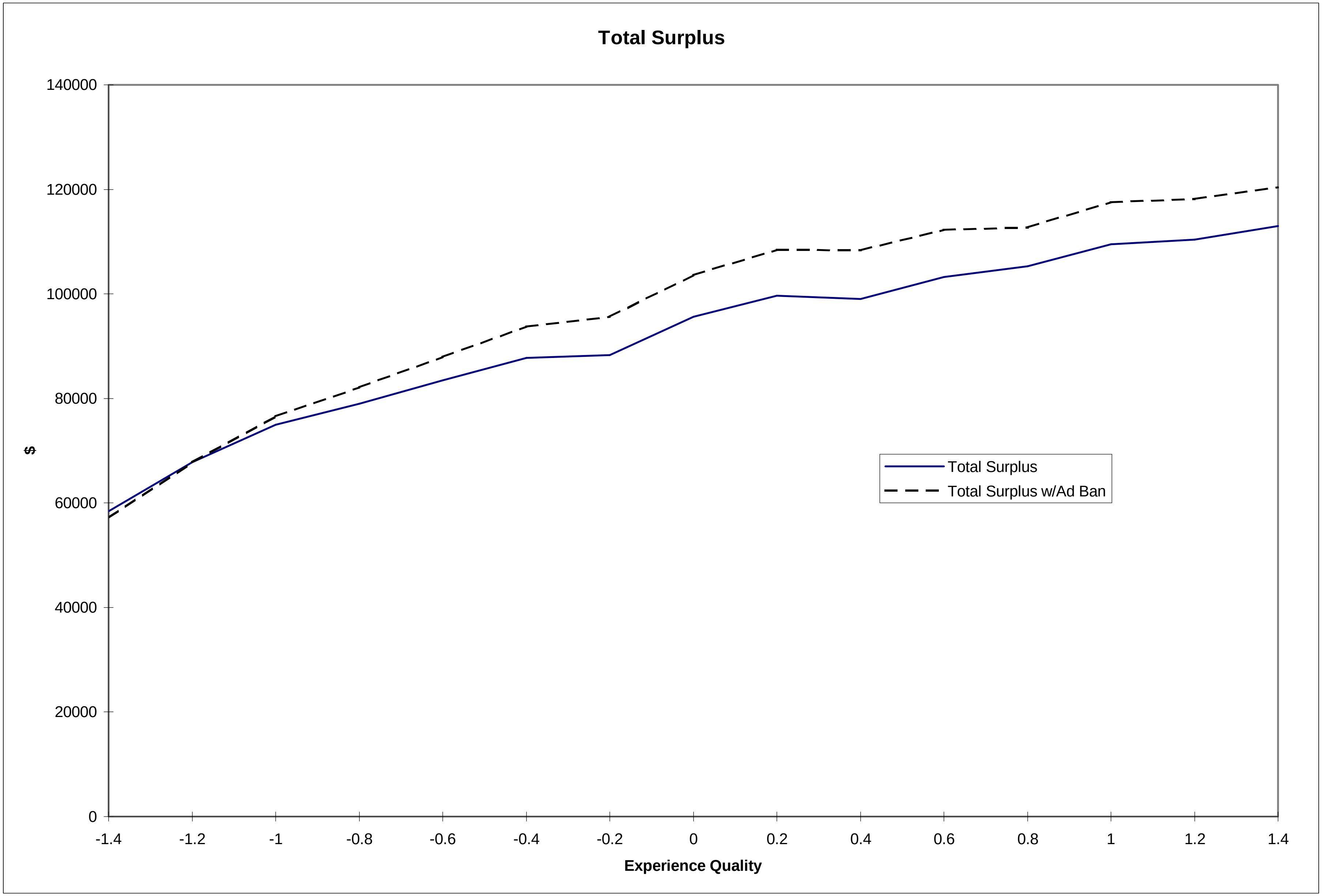


Figure 5.7



Total Surplus
140000
120000
100000
80000
60000
40000
20000
0
$
-1.4
-1.2
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
1.2
1.4
Experience Quality
Total Surplus
Total Surplus w/Ad Ban

References

Ackerberg D. (1996) "Empirically Distinguishing Informative and Prestige Effects of Advertising" mimeo, Boston University

Ackerberg, D. (1996b) "Advertising, Information, and Consumer Choice in Experience Good Markets" Doctoral Dissertation, Yale University.

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

Becker, G.S. and Murphy, K.M. (1993) "A Simple Theory of Advertising as a Good or Bad", *Quarterly Journal of Economics*, 942-964

Benham, L. (1972) "The Effect of Advertising on the Price of Eyeglasses" *Journal of Law and Economics* 337-352

Berry, S. , Levinsohn, J., and Pakes, A. (1995) "Automobile Prices in Market Equilibrium" *Econometrica.,* 63, 841-890

Boyer, K.D. (1974) "Informative and Goodwill Advertising" *The Review of Economics and Statistics* 541-548

Butters, G. (1977) "Equilibrium Distributions of Sales and Advertising Prices" *Review of Economic Studies* 44:465-491

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

Deighton, J., Henderson, C. and Neslin, S. (1994) "The Effects of Advertising on Brand Switching and Repeat Purchasing" *Journal of Marketing Research*, 31:28-41.

Dixit, A. and Norman, V. (1978) "Advertising and Welfare" *Bell Journal of Economics* 9:1-17

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

Eckstein, Z., Horsky, D., and Raban, Y. (1988) "An Empirical Dynamic Model of Optimal Brand Choice", Mimeo, Tel-Aviv University.

Eckstein, Z. and Wolpin, K. (1991) "The Specification and Estimation of Dynamic Stochastic Discrete Choice Models" *The Journal of Human Resources*, 24:563-597.

Galbraith, J.K. (1958) *The Affluent Society* Houghton-Mifflin, Boston, MA

Gonul, F. and Srinivasan, K. (1993) "Modeling Multiple Sources of Heterogeneity in Multinomial Logit Models: Methodological and Managerial Issues" *Marketing Science* 12:213-230

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

Guadagni, P. and Little J., (1983) "A Logit Model of Brand Choice Calibrated on Scanner Data" *Marketing Science*, 2:206-238

Johnson, S., Stedinger, J., Shoemaker, C., Li, Y. Tejada-Guibert, J. (1993) "Numeric Solution of Continuous State Dynamic Programs Using Linear and Spline Interpolation", *Operations Research*, 41-3 484-500

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

Keane, M. and Wolpin, K. (1994) "The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence" *Review of Economics and Statistics* 648-672

Kihlstrom and Riordan (1984) "Advertising as a Signal" *Journal of Political Economy* 92:427-450

Lancaster, K. (1971) *Consumer Demand: A New Approach* New York, Columbia University Press

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

Marshall, A. (1919) *Industry and Trade; a Study of Industrial Technique and Business Organization* London, MacMillan

McCulloch, R. and Rossi, P. (1994) "An Exact Likelihood Analysis of the Multinomial Probit Model" *Journal of Econometrics* 64:207-240

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

McFadden, D. (1989) "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration" *Econometrica* 57:995-1026

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

Nelson, P (1970) "Information and Consumer Behavior", *Journal of Political Economy* 78:311-329

Nelson, P. (1974) "Advertising as Information" *Journal of Political Economy* 82:729-753

Pakes, A. (1986) "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks" *Econometrica* 54:755-85

Pakes, A. and Pollard, D. (1989) "Simulation and the Asymptotics of Optimization Estimators" *Econometrica* 57:1027-57

Pedrick, J. aand Zufryden, F. (1991) "Evaluating the Impact of Advertising Media Plans: A Model of Consumer Purchase Dynamics using Single Source Data" *Marketing Science* 10:111-129

Resnick, A. and Stern, B. "An Analysis of the Information Content in Television Advertising" *Journal of Marketing* 41:50-53

Russell, G. and Kamakura, W. (1994) "Understanding Brand Competition Using Micro and Macro Scanner Data" *Journal of Marketing Research* 31:289-307

Rust, J. (1987) "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher" *Econometrica* 55:999-1035

Rust, J. (1995) "Numerical Dynamic Programming in Economics" Draft for *Handbook of Computational Economics*

Rust, J. (1997) "Using Randomization to Break the Curse of Dimensionality" Forthcoming *Econometrica*

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

Schmalensee R. (1978) "A Model of Advertising and Product Quality" *Journal of Political Economy*, 86

Shapiro, Carl (1982) "Consumer Information, Product Quality, and Seller Reputation" *Bell Journal of Economics*, 20-35

Stigler, G.L. (1961) "The Economics of Information" *Journal of Political Economy* 71:213-225

Stigler, G and Becker, G. (1977) "De Gustibus Non Est Disputandum" *American Economic Review*, 471-99

Telser, L.G. (1964) "Advertising and Competition", *Journal of Political Economy,* 72:537-562

Tirole, J. (1988) *The Theory of Industrial Organization* Cambridge, MA, MIT Press