Discrete Choice Models with Multiple Unobserved Choice Characteristics*

Susan Athey† and Guido W. Imbens‡

April 2007

Abstract

Since the pioneering work by Daniel McFadden in the 1970s and 1980s (McFadden, 1973, 1981, 1982, 1984) discrete (multinomial) response models based on utility maximization have become an important tool of empirical researchers. A key feature of these models is the specification of utilities associated with the alternatives in terms of choice characteristics and individual preferences. Various generalizations of the basic models have been developed to allow for heterogeneity in taste parameters and heterogeneity in product characteristics that is unobserved to the econometrician. In this paper we investigate how rich a specification of the unobserved components is needed to rationalize arbitrary patterns of choice data (generated by utility-maximizing behavior) in settings with many individual decision makers and large choice sets. We find that in general the model must include at least one unobserved choice characteristic. If, as in common, one restricts the utility function to be monotone in the unobserved choice characteristic, then up to two unobserved choice characteristics may be needed to rationalize the choice data. We illustrate the results using scanner data about yoghurt purchases, employing a Bayesian estimation strategy that is particularly well suited to dealing with multiple unobserved product characteristics. We find that the inclusion of two unobserved choice characteristics leads to more reasonable estimates of elasticities.

JEL Classification: C14, C21, C52

Keywords: *Discrete Choice Models, Hedonic Models, Bayesian Methods, Markov Chain Monte Carlo, Structural Models*

*This paper was prepared for the conference "Economics to Econometrics Conference," held in May 2005 in honor of Daniel McFadden. We are grateful to the Dan McFadden, Charles Manski, other conference participants and an anonymous reviewer for comments. Financial support for this research was generously provided through NSF grants SES 0351500 (Athey) and SES 0452590 (Imbens).

†Department of Economics, Harvard University, Cambridge, MA 02138, and NBER. Electronic correspondence: athey@fas.harvard.edu, http://www.economics.harvard.edu/faculty/athey/athey.html.

‡Department of Economics, Harvard University, Cambridge, MA 02138, and NBER. Electronic correspondence: imbens@harvard.edu, http://www.economics.harvard.edu/faculty/imbens/imbens.html.

1 Introduction

Since the pioneering work by Daniel McFadden in the 1970s and 1980s (McFadden, 1973, 1981, 1982, 1984; Hausman and McFadden, 1984) discrete (multinomial) response models have become an important tool of empirical researchers. They remain one of the leading examples of the benefits of a tight integration of economic theory and econometrics. McFadden's early worked focused the application of logit-based choice models to transportation choices. Since then these models have been applied in many areas of economics, including labor economics, public finance, development, finance, and others. One of the currently most active area of applications of these methods is to demand analysis for differentiated products in industrial organization. A common feature of these applications is the presence of many choices.

The application of McFadden's methods to industrial organization has inspired numerous extensions and generalizations of the basic multinomial logit model. As pointed out by McFadden, multinomial logit models have the Independence of Irrelevant Alternatives (IIA) property, so that, for example, an increase in the price for one good implies a redistribution of part of the demand for that good to the other goods in proportions equal to their original market shares. This places strong restrictions on the substitution patterns (cross-price elasticities) of products: elasticities are proportional to market shares. McFadden proposed various extensions to the standard model in order to relax the IIA property and generate more realistic substitution patterns, including "nested logit" models and "mixed logit" models. The subsequent literature has explored extensions to and implementations of these ideas. The nested logit model allows for layers of choices, grouped into a tree structure, where within a nest the IIA property is imposed, but not across nests (McFadden, 1982; Goldberg, 1995; Bresnahan, Trajtenberg, and Stern 1997). The random coefficients or mixed logit approach was generalized in an influential pair of papers by Berry, Levinsohn and Pakes (1995, 2004) (BLP from hereon) and applied to settings with a large number of choices. BLP developed methods for estimating models with random coefficients on product attributes (mixed logit models) as well as unobserved choice characteristics in settings with aggregate data. Exploiting the logistic structure of the model, Berry (1994) proposed a method to relate market shares to a scalar unobserved choice characteristic. BLP also introduced computational tools, building on the simulation methods proposed by McFadden (1989) and Pakes and Pollard (1989), to make these models tractable and showed that they are sufficiently flexible to generate realistic substitution patterns. Their methods have found widespread application.

The hedonic approach, where the utility is modeled as a parametric function of a finite number of choice characteristics and a finite number of individual characteristics, has recently attracted renewed interest. Researchers have considered hedonic models both with and without individual-choice specific error terms (Bajari and Benkard, 2004; Berry and Pakes, 2001). These models have some attractive properties, especially in settings with many choices, because the number of parameters does not increase with the number of choices. Unlike the nested and random coefficient logit models, they do not imply that all choices will end up with positive market shares. On the other hand, simple forms of those models rule out particular choices for individuals with specific characteristics, making them very sensitive to misspecification. To make these models more flexible researchers have typically allowed for unobserved choice and individual characteristics. To maintain computational feasibility the number of unobserved choice characteristics is typically limited to one.

This paper explores a version of the multinomial choice model that has received less attention in the literature. We consider a random coefficients model of individual utility that includes observed individual and product characteristics, as well as multiple unobserved product characteristics and unobserved individual preferences for both observed and unobserved product characteristics. The idea of specifying such a model goes back at least to McFadden (1981), but only a few papers have followed this approach (e.g. Elrod and Keane, 1995, Harris and Keane, 1999; Keane, 1997, 2004; Goettler and Shachar, 2001). This model has several desirable features. For example, the model nests both models based on unobserved product characteristics (BLP) as well as unrestricted multinomial probit models (e.g. McCulloch, Polson and Rossi (2000) (MPR)). In addition, by describing products as combinations of attributes, it is possible to consider questions about the introduction of new products in particular parts of the product space.

In many cases researchers applying this class of models have employed restrictions on the number of unobserved choice characteristics. In other cases (e.g. Goettler and Shachar (2001)) authors have allowed for a large number of choice characteristics, with the data determining the number of unobserved characteristics that enter the utility function. However, the literature has not directly considered the question of what restrictions are implied by limiting the number of choice characteristics, nor is it clear whether, in the absence of parametric restrictions, the data can provide evidence for the existence of multiple unobserved product characteristics. Understanding the answers to these questions is important for empirical researchers who may not always be aware of the implications of the modeling choices. Although researchers may still find it useful to apply a model that cannot rationalize all patterns of choice data, we argue that the researcher should be aware of any limitations the model imposes in this regard.

In this paper, we provide formal results to address these questions. We begin by asking how flexible a model is required – that is, how many and what kind of unobserved variables must be included in the specification of consumer utility– to rationalize choice data. We are interested in whether any pattern of market shares that might be consistent with utility maximization can be rationalized. We discuss settings and data configurations where one can establish that the utility function must depend on multiple unobserved choice characteristics rather than a single unobserved product characteristic. We also discuss the extent to which models with no unobserved individual characteristics can rationalize observed data.

We explore the implications of these models in an application to demand for yogurt. We consider models with up to two unobserved choice characteristics, and assess the implied price elasticities. In order to implement these models we employ Bayesian methods. Such methods have been used extensively in multinomial choice settings by Rossi, McCullach and Allenby (1996), MPR, McCulloch and Rossi (1994), Allenby, Chen, and Yang (2003), Rossi, Allenby and McCulloch (2005), Bajari and Benkard (2003), Chib and Greenberg (1998), Geweke and Keane (2002), Romeo (2003), Osborne (2005) and others. These authors have argued that Bayesian methods are very convenient for latent index discrete choice models with large numbers of choices, using modern computational methods for Bayesian inference, in particular data augmentation and Markov- Chain-Monte-Carlo (MCMC) methods (Tanner and Wong, 1987; Chib, 2003; Geweke, 1997; Gelman, Carlin, Stern and Rubin, 2004; Rossi, Allenby and McCulloch, 2005). See Train (2003) for a comparison with frequentist simulation methods.

2 The Model

Consider a model with M "markets," where markets might be distinguished by variation in time as well as location. In market m there are N_m consumers, each choosing one product from a set of J products.[1] In this market product j has two sets of characteristics, one observed, denoted by X_{jm}, and one unobserved, denoted by ξ_j. The observed product characteristics may vary by market, though they need not do so. The vector of unobserved product characteristics does not vary by market.[2] The vector of observed product characteristics X_{jm} is of dimension K, and the vector of unobserved product characteristics ξ_j is of dimension P. Individual i has a vector of observed characteristics Z_i (which for notational convenience includes a constant term) of dimension L, and a vector of unobserved characteristics ν_i of dimension $K + P$.[3]

The utility associated with choice j for individual i in market m is U_{ijm}, for $i = 1, \ldots, N_m$, $j = 1, \ldots, J$, and $m = 1, \ldots, M$. Individuals choose product j if the associated utility is higher than that associated with any of the alternatives.[4] Hence the probability that an individual in market m with characteristics z chooses product j is

$$s_{jm}(z) = \Pr(U_{ijm} > U_{ikm} \text{ for all } k \neq j | \, X_{1m}, \ldots, X_{Jm}, Z_i = z). \tag{2.1}$$

We assume there is a continuum of consumers in each market so that this probability is equal to the market share for product j in market m among the subpopulation with characteristics z.

We consider the following model for U_{ijm}:

$$U_{ijm} = g(X_{jm}, \xi_j, Z_i, \nu_i) + \epsilon_{ijm},$$

[1] In the implementation we allow for the possibility that in some markets only a subset of the products is available. In order to keep the notation simple we do not make this explicit in the discussion in this section. Similarly, we allow for multiple purchases by the same individual, although the notation does not make this explicit at this point.

[2] We make the assumption that unobserved product characteristics do not vary by market a defining characteristic of multiple markets with the same goods (conditional on observables): if products vary across markets in unobservable ways, there is little value to having observations from multiple markets absent additional assumptions about the way in which these unobservables vary across markets. Some authors (e.g. Petrin and Train, 2005) have argued that since in equilibrium, prices respond to unobservable product characteristics, prices are informative about these characteristics. Their approach relies on equilibrium pricing assumptions, which are clearly more appropriate in some settings than in others (e.g. regulated markets). We do not pursue that approach here.

[3] We assume that the dimension of the unobserved individual component is equal to the sum of the number of observed and unobserved choice characteristics, allowing each choice characteristic to have its own individual-specific effect on utility. Although we do establish the importance of allowing for unobserved individual heterogeneity, we do not explore the extent of this need. It may not be necessary to allow the dimension of the unobserved individual heterogeneity to be as large as $K + P$.

[4] We ignore the possibility of ties in the latent utilities. In the specific models we consider such ties would occur with probability zero.

where g is unrestricted, and the additional component ϵ_{ijm} is assumed to be independent of observed and unobserved product characteristics and observed and unobserved individual characteristics. It is also assumed to be independent accross choices, markets and individuals, and have a logistic distribution. This idiosyncratic error term is interpreted as incorporating individual-specific preferences for a product that are unrelated to all other product features.

Let us briefly consider a parametric version of this model in order to relate it more closely to models used in the empirical literature. Suppose the systematic part of the utility has the form

$$g(X_{jm}, \xi_j, Z_i, \nu_i) = X_{jm}'\beta_i + \xi_j'\gamma_i,$$

where the individual specific marginal utilities β_i and γ_i relate to the observed and unobserved individual characteristics through the equation:

$$\begin{pmatrix} \beta_i \\ \gamma_i \end{pmatrix} = \begin{pmatrix} \Delta_o \\ \Delta_u \end{pmatrix} Z_i + \begin{pmatrix} \nu_{oi} \\ \nu_{ui} \end{pmatrix} = \Delta Z_i + \nu_i.$$

In this representation β_i is a K-dimensional column vector, γ_i is an P-dimensional column vector, Δ is a $(K+P)\times L$-dimensional matrix, and ν_i is a $(K+P)$-dimensional column vectors. The unobserved components of the individual characteristics are assumed to have a normal distribution:

$$\nu_i | \mathbf{X}_m, Z_i \sim \mathcal{N}(0, \Omega),$$

where $\mathbf{X}_m$ is the $J \times K$ matrix with jth row equal to X_{jm}', and Ω is a $(K+P)\times(K+P)$-dimensional matrix. Now we can write the utility as

$$U_{ijm} = X_{jm}'\Delta_o Z_i + \xi_j'\Delta_u Z_i + X_{jm}'\nu_{oi} + \xi_j'\nu_{ui} + \epsilon_{ijm}. \tag{2.2}$$

We contrast this model with three models that have been discussed and used more widely in the literature. The first is the special case with no unobserved product or individual characteristics:

$$U_{ijm} = X_{jm}'\Delta_o Z_i + \epsilon_{ijm}.$$

This is the standard multinomial logit model (McFadden, 1973). It has the IIA property that the conditional probablity of making choice j rather than k, given that one of the two is choosen, does not depend on characteristics of other choices. This in turn implies severe restrictions on cross-elasticities and thus on substitution patterns. For a general discussion, see McFadden (1982, 1984).

A second alternative model features a single unobserved product characteristic and unobserved individual characteristics.

$$U_{ijm} = X_{jm}'\beta_i + \xi_j + \varepsilon_{ij} = X_{jm}'\Delta_o Z_i + \xi_j + X_{jm}'\nu_{oi} + \epsilon_{ijm}.$$

This is a special case of the model used in BLP (who allow for endogeneity of some of the observed product characteristics, which for simplicity we do not consider here). This model allows for much richer patterns of substitution, while remaining computationally tractable even in settings with many choices. This model, with the generalization to allow for endogeneity of some choice characteristics, has become very popular in the applied literature. See Ackerberg, Benkard, Berry and Pakes (2006) for a recent survey.

The third model is typically set up in a different way. Suppose

$$U_{ijm} = X_{jm}' \Delta Z_i + \varepsilon_{ijm},$$

with unrestricted dependence between the unobserved components for different choices,

$$\begin{pmatrix} \varepsilon_{i1m} \\ \varepsilon_{i2m} \\ \vdots \\ \varepsilon_{iJm} \end{pmatrix} \sim \mathcal{N}(0, \Omega),$$

where $\varepsilon_{i\cdot m}$ is the J vector with all ε_{ijm} for individual i in market m, with the $J \times J$ matrix Ω not restricted (beyond some normalizations). This is the type of model studied in MPR and McCulloch and Rossi (1994). We can relate this to the set up in (2.2) as follows. Let

$$U_{ijm} = X_{jm}' \Delta_o Z_i + \xi_j' \nu_{ui},$$

with the dimension of the vector of unobserved choice characteristics ξ_j and the vector of unobserved individual characteristics ν_{ui} both equal to J. Moreover, suppose that all elements of the J-vector ξ_j are equal to zero other than the j-th element which is equal to one. Then $\epsilon_{ijm} = \xi_j' \nu_{ui} = \nu_{uij}$ and

$$\begin{pmatrix} \varepsilon_{i1m} \\ \varepsilon_{i2m} \\ \vdots \\ \varepsilon_{iJm} \end{pmatrix} = \begin{pmatrix} \xi_1 & \xi_2 & \dots & \xi_J \end{pmatrix}' \nu_{ui} = \nu_{ui} \sim \mathcal{N}(0, \Omega).$$

The insight from this representation is that we can view the MPR set up as equivalent to (2.2) by allowing for as many unobserved choice characteristics as there are choices. The view underlying this approach is that choices are fundamentally different in ways that cannot be captured by a few characteristics.

Our discussion below will focus largely on the need for unobserved choice characteristics in order to explain data on choices arising from utility maximizing individuals. We will argue that in the absence of functional form restrictions a single unobserved product characteristic as in the BLP set up may not suffice to rationalize all choice data, but that the MPR approach allows for more unobserved choice characteristics than the data can ever reveal the existence of. We show that two unobserved choice characteristics are sufficient, even in the case with many choices, to rationalize choice data arising from utility maximizing behavior. By providing formal support for the ability of characteristic-based models to rationalize choice data, this discussion complements the substantive discussion in, among others, Ackerberg, Benkard, Berry and Pakes (2006) who argue in favor of characteristics-based approaches, and and the contrasting arguments in Kim, Allenby, and Rossi (forthcoming), who argue in favor of the view that generally choices cannot be captured by a low-dimensional set of characteristics.

3 Some Results on Rationalizability of Choice Data

In Section 2, we introduced a general nonparametric model. In this section we consider the ability of this model to rationalize data arising from choices based on utility maxizing behavior. Our model decomposes individual-product unobservables into individual observed and unobserved preferences (random coefficients) for observed and unobserved product characteristics, where individual- and product-level unobservables interact. An initial question concerns how different types of variation that might be present in a dataset potentially shed light on the importance of various elements of the model. In particular, we ask whether the data can in principle reject restricted versions of the model, such as a model with a single unobserved product characteristic, or a model with homogeneous individuals conditional on observables.

A model is said to be testable if it cannot rationalize all hypothetical datasets that might be observed. Questions about identification and testability are generally considered in the context of hypothetical data sets that are large in some dimension. Typically we consider settings with independent draws from a common distribution, and the limit is based on the number of draws going to infinity. In the current setting, there are several different dimensions where the data set may be large. Specifically, we will consider settings with a large number of individuals facing the same choice set (large N_m), where each choice corresponds to a vector of characteristics. We will also consider settings where the number of choices or products itself is large (large J), so that for each product there is a nearby product (in terms of observed product characteristics). Such settings have been the motivation for BLP and literature that follows them (e.g., Nevo, 2000, 2001; Petrin, 2003; Ackerberg and Rysman, 2002; Bajari and Benkard, 2003). Finally, we will consider a large number of markets (large M), where some observed choice characteristics may vary between markets (but all unobserved choice characteristics are constant within markets).

We shall see that a data set with a large number of choices can be used to distinguish between the absence or presence of unobserved choice characteristics, and that a data set with a large number of markets and sufficient variation in observed product characteristics can be used to establish the presence of unobserved individual heterogeneity.

3.1 Rationalizability in a Single Market

In this subsection, we set $M = 1$ and suppress the subscript indicating the market in our notation. First consider the case with a finite number of choices J and a large number of individuals. We can summarize what we can learn from the data in terms of the conditional probability of choice j given individual characteristics $Z_i = z$. Denote this probability, equal to the market share because we have a large number of individuals in each market, by $s_j(z)$. Initially consider a simple setting with no unobserved individual and no unobserved choice characteristics. Let the utility associated with choice j for individual i be $U_{ij} = g(X_j, Z_i)$, without functional form assumptions. Consider the subpopulation with characteristics $Z_i = z$. Within this subpopulation all individuals face the same decision problem,

$$\max_{j \in \{1,\ldots,J\}} g(X_j, z).$$

Since we have no randomness in this simplified model, the market shares $s_j(z)$ implied by this model are degenerate: if individual i with characteristics $Z_i = z$ prefers product j, then $g(X_j, z) > g(X_k, z)$ for all $k \neq j$, so that any other individual i' with $Z_{i'} = z$ would make the same choice. Hence, under this model we would expect to see a degenerate distribution of choices conditional on the individual characteristics. Specifically, all individuals would choose j, where $j = \arg\max_{j'=1,\ldots,J} g(X_{j'}, z)$, so that for this j we have $s_j(z) = 1$, and for all other choices $k \neq j$ we would see $s_k(z) = 0$. Hence, as soon as we see two individual with the same observed individual characteristics making different chocies, we can reject such a model with certainty.

Now suppose that in addition to the observed choice and individual characteristics there is an additive idiosyncratic error term ϵ_{ij}, independent across choices and individuals. The utility associated with individual i and choice j is then $g(X_j, Z_i) + \epsilon_{ij}$. In that case we would see a distribution of choices even within a subpopulation homogenous in terms of the observed individual characteristics, and we would see $s_j(z) > 0$ for all $j = 1, \ldots, J$ given large enough support for ϵ_{ij}.

For purposes of exposition, suppose that the ϵ_{ij} have an extreme value distribution (although for computational reason we will consider normally distributed ϵ_{ij} when implementing the model from Section 5.1). Then the probabilities $s_j(z)$ have a logit form:

$$s_j(z) = \frac{\exp(g(x_j, z))}{\sum_{k=1}^{J} \exp(g(x_k, z))}.$$

This in turn implies that the log of the ratio of the probability of choice j versus choice k has the form

$$\ln\left(\frac{s_j(z)}{s_k(z)}\right) = g(X_j, z) - g(X_k, z).$$

We can normalize the functions $g(x, z)$ by setting $g(X_1, z) = 0$. For a finite number of choices, all with unique characteristics, we can always find a continuous function $g(x, z)$ that satisfies this restriction for all pairs (j, k). Hence in this setting we cannot reject the semi-parametric version of the conditional logit model, nor its implication of independence of irrelevant alternatives.

One reason we cannot reject the simple model is that we do not see individuals choosing among products that appear similar. In other words, there need not be choices with similar observable characteristics. In order to investigate identification issues further it is useful to consider a setting with a large number of choices where some choices would be similar in observable characteristics.

Following Berry, Linton and Pakes (2003), suppose that for all choices j and for all individual characteristics z the choice probabilities, normalized by the number of choices J, are bounded away from zero and one, so that $0 < \underline{c} \leq J \cdot s_j(z) \leq \overline{c} < 1$. Suppose that we observe $J \cdot s_j(z)$ for a large number of choices and all $z \in \mathbb{Z}$. With the choice characteristics in a compact subset of $\mathbb{R}^K$, it follows that eventually we will see choices with very similar observed characteristics. Now suppose we have two choices j and k with X_j equal to X_k. In that case we should see identical choice probability within a given subpopulation, or $s_j(z) = s_k(z)$. If in fact we find that the choice probabilities differ, the model is misspecified. One possible source of misspecification is an unobserved choice characteristic. Note that the finding $s_j(z) \neq s_k(z)$ can *not* be explained by (unobserved) heterogeneity in individual preferences: if the two products are identical in all characteristics, their market shares within the same market should be identical (given the additional independent error ϵ_{ij}).

Now let us consider whether, and under what conditions, it is sufficient to have a single unobserved product characteristic. Much of the existing literature (e.g. BLP) assumes that the utility function is strictly monotone in the unobserved choice characteristics for each individual, and that there is a single unobserved product characteristic. We now argue that this combination of assumptions can be rejected by the data. Without loss of generality assume that $g(x, z, \xi)$ is nondecreasing in the scalar unobserved component ξ. Consider two choices j and k with the same values for the observed choice characteristics, $X_j = X_k$. Suppose that for a given subpopulation with observed characteristics $Z_i = z$ we find that $s_j(z) > s_k(z)$. We can infer that the unobserved choice characteristic for product j is larger than that for product k: $\xi_j > \xi_k$. Now suppose we have a second subpopulation with different individual characteristics $Z_i = z'$. The assumption of monotonicity of the utility function in ξ implies that the same ordering of the choice probabilities must hold for this second subpopulation: $s_j(z') > s_k(z')$. If we find that $s_j(z') < s_k(z')$, the original model must be misspecified.

One possible source of misspecification is the presence of multiple unobserved choice characteristics. Suppose there are two unobserved choice characteristics ξ_{j1} and ξ_{j2}. In that case it could be that individuals with $Z_i = z$ put more weight in the utility function on the first characteristic $\xi_{\cdot 1}$, and as a result prefer product j to product k because $\xi_{j1} > \xi_{k1}$, and individuals with $Z_i = z'$ put more weight on the second characteristic $\xi_{\cdot 2}$ and prefer product k to j because $\xi_{j2} < \xi_{k2}$. This argument shows that in settings with a single market and no variation in product characteristics, the presence of multiple choices with similar observed choice characteristics can imply the presence of at least two choice characteristics under monotonicity of the utility function in the unobserved choice characteristic. Again, the presence of unobserved individual heterogeneity cannot explain the pattern of the probabilities described above. An alternative source of misspecification has been considered in an interesting study of the demand for television shows by Goetler and Shachar (2001). They allow for the presence of multiple unobserved characteristics that enter the utility function in a non-monotone manner (in their application consumers have a bliss point in each unobserved choice characteristics, and utility is quadratic; each consumer's bliss point is unrestricted). Models with multiple unobserved product characteristics have been considered in an interesting series of papers by Keane and coauthors (Elrod and Keane, 1995; Harris and Keane, 1997; Keane, 1997, 2004), and in work by Poole and Rosenthal (1985).

Here, we argue that with a flexible model and a countable number of products, a single dimension of unobserved product characteristics can rationalize the data. However, it is necessary that utility be nonmonotone in this unobservable. With a restriction to utility that is monotone in the unobservable, it is not sufficient to have a single unobserved product characteristic. However, one can say more. In the example it was possible to rationalize the data with two unobserved choice characteristics that enter the utility function monotonically. We show that this is true in general. The following theorem formalizes this. The setting is one with a countable number of products with identical observed product characteristics, and a compact set of observed individual characteristics. There are many individuals, so the market shares $s_j(z)$ are known for all $z \in \mathbb{Z}$ and for all $j = 1, \ldots, J$. We show that irrespective of the number of products J we can rationalize the pattern of market shares with a utility function that is increasing in two unobserved product characteristics.

Theorem 3.1 *Suppose that for each subpopulation indexed by characteristics $z \in \mathbb{Z}$, and for all $J = 1, \ldots, \infty$, there exist J products with identical observed characteristics and an observable vector of market shares $s_{jJ}(z)$, $j = 1, \ldots, J$, such that $\sum_{j=0}^{J} s_{jJ}(z) = 1$. Then we can rationalize these market shares with a utility function*

$$U_{ij} = g(Z_i, \xi_j) + \epsilon_{ij},$$

where ξ_j is a scalar, ϵ_{ij} has an extreme value distribution and is independent of ξ_j, and where $g(z, \xi)$ continuous in ξ. Moreover we can also rationalize these market shares with a utility function

$$U_{ij} = h(Z_i, \xi_{1j}, \xi_{2j}) + \epsilon_{ij},$$

where ξ_{1j}, ξ_{2j} are scalars, ϵ_{ij} has an extreme value distribution and is independent of ξ_{1j}, ξ_{2j}, and where $h(z, \xi_1, \xi_2)$ continuous and monotone in ξ_1 and ξ_2.

Proof: The proof is constructive. Under the assumptions in the theorem we can infer the market shares $s_j(z)$ for all choices and all values of z. The form of the utility function implies that the market shares have the form

$$s_j(z) = \frac{\exp(g(z, \xi_j))}{\sum_{k=1}^{J} \exp(g(z, \xi_k))}.$$

Define $r_j(z) = \ln\left(s_j(z)/s_1(z)\right)$ (so that $r_1(z) = 0$). The proof of the first part of the theorem amounts to constructing a function $g(z, \xi)$ and a sequence $\xi_1, \ldots, \xi_J$ such that $r_j(z) = g(z, \xi_j)$ for all z and j. First, let

$$\xi_j = 1 - 2^{-j}, \text{for } j = 1, \ldots, J. \tag{3.3}$$

Next, for $\xi \in [0, 1]$

$$g(z, \xi) = \begin{cases} r_j(z) & \text{if } \xi = 1 - 2^{-j}, j = 1, \ldots, J \\ 0 & \text{if } 0 \leq \xi < 2^{-1} \\ r_j(z) + \frac{\xi - (1-2^{-j})}{2^{-j} - 2^{-(j+1)}} \cdot (r_{j+1}(z) - r_j(z)) & \text{if } 1 - 2^{-j} < \xi < 1 - 2^{-(j+1)} \\ r_J(z) & \text{if } 1 - 2^{-J} < \xi \leq 1. \end{cases} \tag{3.4}$$

This function $g(z, \xi)$ is continuous in ξ on $[0, 1]$ for all z, and piece-wise linear with knots at $1 - 2^{-j}$.

To construct the function $h(z, \xi_1, \xi_2)$ we use the fact that a continuous function $k(\xi)$ of bounded variation on a compact set can be written as the sum of a nondecreasing continuous function $k_1(\xi)$ and a nonincreasing function $k_2(\xi)$. We apply this to the function $g(z, \xi)$ in (3.4) for each value of z so that $g(z, \xi) = h_1(z, \xi) + h_2(z, \xi)$ with $h_1(z, \xi)$ nondecreasing and $h_2(z, \xi)$ nonincreasing, and both continuous. Then define

$$h(z, \xi_1, \xi_2) = h_1(z, \xi_1) + h_2(z, 1 - \xi_2), \tag{3.5}$$

which is by construction nondecreasing and continuous in both ξ_1 and ξ_2. Then choose $\xi_{1j} = \xi_j$ and $\xi_{2j} = 1 - \xi_j$, where ξ_j is as defined in equation (3.3), and the function satisfies

$$h(z, \xi_{1j}, \xi_{2j}) = h(z, \xi_j, 1 - \xi_j) = h_1(z, \xi_j) + h_2(z, \xi_j) = g(z, \xi_j) = r_j(z). \tag{3.6}$$

□

In both cases, utility will potentially be highly nonlinear in the unobservable, and so with a restriction to linear and monotone effects of the unobservables, a particular functional form might fit better with multiple dimensions of unobservables, to capture nonlinearities in the true model.

The restriction in the theorem that all products have the same observed characteristics is imposed only to simplify the notation. We can allow for a finite set of different values for the observed product characteristics. More generally, we interpet this theorem as demonstrating that unless one allows for utility functions that are highly nonlinear, with derivatives large in absolute value, one may need two unobserved product characteristics (or one if one allows for non-monotonicity in this unobserved product characteristic), in order to rationalize arbitrary patterns of market shares.

This result does not imply that the function $h(z, \xi_{1j}, \xi_{2j})$ is identified, even after normalization restrictions have been imposed. There may be many different such functions that rationalize the data. Establishing what additional assumptions and normalizations are required for identification, particularly for models that include unobserved individual heterogeneity, remains an open problem.

3.2 Rationalizability in Multiple Markets

In this subsection we consider the evidence for the presence of unobserved heterogeneity at the individual level. To some extent allowing for such heterogeneity substitutes for heterogeneity in unobserved choice characteristics. It was argued before that in the case with no unobserved choice or individual characteristics one would expect to see the choice probabilities be equal to zero or one. Introducing unobserved individual characteristics will generate a distribution of choices in that case. More importantly, however, is that unobserved choice characteristics generate substitution patterns that are more realistic. Consider again a situation with a large number of individuals and a finite number of choices J. We have already argued that such a model fits the data arbitrary well. However, suppose that we have data from multiple markets. Markets may be distinguished by geography or time. These markets have different populations, and thus potentially different distributions of individual characteristics. We assume that the choice set is the same in all markets, but the observed choice characteristics of the products may differ between markets. Key examples of such choice characteristics that vary by market include prices and marketing variables.

In order to discuss this setting we need to return to the general notation of Section 2. Let $m = 1, \ldots, M$ index the markets. In market m there are N_m individuals. They choose between J products, where product j has observed characteristics X_{jm} and unobserved characteristics ξ_j. The general form for the utility for individual i in market m associated with product j is

$$U_{ijm} = g(X_{jm}, \xi_j, Z_i, \nu_i) + \epsilon_{ijm},$$

for $i = 1, \ldots, N_m$, $j = 1, \ldots, J$, and $m = 1, \ldots, M$. The idiosyncratic error ϵ_{ijm} is independent of $\epsilon_{i'j'm'}$ unless $(i, j, m) = (i', j', m')$, and has an extreme value distribution.

First consider a model with no unobserved individual characteristics, so that

$$U_{ijm} = g(X_{jm}, \xi_j, Z_i) + \epsilon_{ijm}.$$

Recall that the unobserved choice characteristics do not vary by market. Consider a subpopulation of individuals with observed characteristics $Z_i = z$. Consider two markets m and m', and three choices, j, k, and l, where for two of the choices, j and k, the characteristics do not differ between markets, and for the third choice, l, the observed characteristics do differ between markets, so that $X_{jm} = X_{jm'}$, $X_{km} = X_{km'}$, and $X_{lm} \neq X_{lm'}$. In this case the market share of choice j in markets m and m' is

$$s_{jm}(z) = \frac{\exp(g(X_{jm}, \xi_j, z))}{\exp(g(X_{jm}, \xi_j, z)) + \exp(g(X_{km}, \xi_k, z)) + \exp(g(X_{lm}, \xi_l, z))},$$

and

$$s_{jm'}(z) = \frac{\exp(g(X_{jm'}, \xi_j, z))}{\exp(g(X_{jm'}, \xi_j, z)) + \exp(g(X_{km'}, \xi_k, z)) + \exp(g(X_{lm'}, \xi_l, z))}.$$

The ratio of the market shares for choices j and k in the two markets are

$$\frac{s_{jm}(z)}{s_{km}(z)} = \frac{\exp(g(X_{jm}, \xi_j, z))}{\exp(g(X_{km}, \xi_k, z))}, \qquad \text{and} \quad \frac{s_{jm'}(z)}{s_{km'}(z)} = \frac{\exp(g(X_{jm'}, \xi_j, z))}{\exp(g(X_{km'}, \xi_k, z))}.$$

These relative market shares are identical in both markets because $X_{jm} = X_{jm'}$ and $X_{km} = X_{km'}$, and the unobserved choice characteristics do not vary by market by assumption. Thus the IIA property of the conditional logit model implies in this case that the ratio of market shares for choices k and j should be the same in the two markets.[5] If the two ratios differ, obviously one possibility is that the unobserved choice characteristics for these choices differ between markets. (Note that a market-invariant choice-specific component would not be able to explain this data pattern.) Ruling this out by assumption, another possibility is that there are unobserved individual characteristics that imply that individuals who are homogenous in terms of observed characteristics do in fact have differential preferences for these choices.

Let us assess how unobserved individual heterogeneity can explain such data. These unobserved individual components are interpreted as individual preferences for product characteristics, such as a taste for quality. As before, let us denote such components by ν_i. For the

[5]Although other functional forms for the distribution of ϵ_{ij} do not impose the independence of irrelevant alternatives property, as long as independence of ϵ_{ij} is maintained, other functional forms also impose testable restrictions on how market shares vary when product characteristics change.

time being we assume the individual unobserved component is independent of observed choice characteristics. The utility becomes

$$U_{ijm} = U(X_{jm}, Z_i, \nu_i) + \epsilon_{ijm},$$

still with the ϵ_{ijm} independent across all dimensions.

4 Predicting the Market Share of New Products

Suppose we wish to predict the market share of a new product, call it choice 0. In order to make such a prediction, the analyst must provide some information about the product's observed and unobserved characteristics. One possibility is to consider products that lie is some specified quantile of the distribution of characteristics in the population. For example, one could consider a product with the median values of observed and unobserved characteristics. However, that may or may not be an interesting hypothetical product to consider, since products in the population may tend to be outliers in some dimensions and not others.

A second alternative approach might be to make some assumptions about the costs of entry and production at various points in the product space, and to calculate the optimal position for a new product. Although an assumption of equilibrium pricing on the part of firms might enable inferences about marginal costs of production for different products, additional assumptions would be required to estimate entry costs at different points.

If there are many products, a third approach would be to model the joint distribution of observed and unobserved product characteristics in the population, and take draws from that joint distribution, thus generating a distribution of predicted market shares. Our estimation routine generates different conditional distributions of unobserved characteristics for each product, and to construct this joint distribution, it would be necessary to combine these estimates with an estimate of the marginal distribution of observed characteristics. Some extrapolation would be required to infer this distribution at values of observed characteristics that are not observed in the population.

Finally, as a fourth approach, in some cases it might be interesting to consider entry of a product with prespecified observed characteristics but unknown unobserved characteristics. For example, a foreign entrant might be planning to introduce an existing product with observable attributes into the markets under study. In that case, the analyst must make some decisions about how to model the unobserved characteristics for this product. One possibility is to use the marginal distribution of unobserved product characteristics in the population. This is the method we use in our empirical application. However, this approach has some important limitations. Most importantly, it does not account for the fact that unobserved characteristics may vary systematically with observed characteristics: for example, prices may vary with unobserved quality. As described in the third approach, it is possible to generate an estimate of the distribution of unobserved characteristics conditional on a particular set of observables, but it requires some extrapolation; since our application has only eight brands, we do not pursue it here.

Following the third or fourth approaches, one immediate implication of the presence of unobserved choice characteristics is that we are unable to predict the market share exactly. Instead, a given set of observable characteristics of a new product would be consistent with a range of market shares. We view this as a realistic feature of the model. Of course, the analyst is free to put more structure on the prediction of the unobservable characteristics, along the lines suggested in the second approach.

5 A Bayesian Approach to Estimation

This section presents a proposed approach for estimating a model with multiple unobserved choice characteristics. Although our rationalizability discussion was largely nonparametric, we focus on estimation of parametric models. Our view is that these can be viewed as approximations to the nonparametric models studied in the previous sections, with our results showing that the evidence for, for example, multiple unobserved product characteristics, is not coming solely from the functional form restrictions. We begin by returning to the parametric model introduced in Section 2, after which we describe a Bayesian approach to estimation. A Bayesian approach is in this case attractive from a computational perspective.

5.1 The Parameterized Model

Recall the general model for U_{ijm}:

$$U_{ijm} = g(X_{jm}, \xi_j, Z_i, \nu_i) + \epsilon_{ijm},$$

where the additional component ϵ_{ijm} is assumed to be independent of $(X_{jm}, \xi_j, Z_i, \nu_i)$. Rather than assume that each ϵ_{ijm} has an extreme value distribution, as we did in some of the discussion above, for the purposes of estimation we assume that it has a standard (mean zero, unit variance) normal distribution, independent of $(X_{jm}, \xi_j, Z_i, \nu_i)$, as well as independent across choices, markets and individuals. We parametrize the systematic part of the utility associated with choice j as

$$g(X_{jm}, \xi_j, Z_i, \nu_i) = X'_{jm}\beta_i + \xi'_j\gamma_i = \begin{pmatrix} X_{jm} \\ \xi_j \end{pmatrix}' \begin{pmatrix} \beta_i \\ \gamma_i \end{pmatrix},$$

where the individual specific coefficients θ_i satisfy:

$$\begin{pmatrix} \beta_i \\ \gamma_i \end{pmatrix} = \begin{pmatrix} \Delta_o \\ \Delta_u \end{pmatrix} Z_i + \begin{pmatrix} \nu_{oi} \\ \nu_{ui} \end{pmatrix} = \Delta Z_i + \nu_i.$$

In this representation β_i is a K-dimensional column vector, γ_i is an P-dimensional column vector, Δ is a $(K+P) \times L$-dimensional matrix, and ν_i is a $(K+P)$-dimensional column vectors. The unobserved components of the individual characteristics are assumed to have a normal distribution:

$$\nu_i | \mathbf{X}_m, Z_i \sim \mathcal{N}(0, \Omega),$$

where $\mathbf{X}_m$ is the $J \times K$ matrix with jth row equal to X'_{jm}, and Ω is a $(K+P) \times (K+P)$-dimensional matrix.

Now we can write U_{ijm} as:

$$\begin{aligned} U_{ijm} &= \begin{pmatrix} X_{jm} \\ \xi_j \end{pmatrix}' (\Delta Z_i + \nu_i) + \epsilon_{ijm} \\ &= X'_{jm}\Delta_o Z_i + \xi'_j \Delta_u Z_i + X'_{jm}\nu_{oi} + \xi'_j \nu_{ui} + \epsilon_{ijm}. \end{aligned}$$

Let us consider the vector of latent utilities for all J choices for individual i in market m:

$$U_{i\cdot m} = \begin{pmatrix} U_{i1m} \\ U_{i2m} \\ \vdots \\ U_{iJm} \end{pmatrix} = \begin{pmatrix} \mathbf{X}_m & \xi \end{pmatrix} \Delta Z_i + \begin{pmatrix} \mathbf{X}_m & \xi \end{pmatrix} \nu_i + \epsilon_{i\cdot m}, \tag{5.7}$$

where ξ is the $J \times P$ matrix with jth row equal to ξ'_j. Conditional on $\mathbf{X}_m$, Z_i, and ξ the joint distribution of the J-vector $U_{i\cdot m}$ is

$$U_{i\cdot m}|\mathbf{X}_m, Z_i, \xi \sim \mathcal{N}\Big(\begin{pmatrix} \mathbf{X}_m & \xi \end{pmatrix} \Delta Z_i, \begin{pmatrix} \mathbf{X}_m & \xi \end{pmatrix} \Omega \begin{pmatrix} \mathbf{X}_m & \xi \end{pmatrix}' + I_J \Big).$$

This model imposes considerable structure on the correlation between the latent utilities, with the covariance matrix and the mean parameters intricately linked, but at the same time does allow for complex patterns in this correlation structure.

5.2 Posterior Calculations

In order to estimate the parameters of interest and do inference we use a Bayesian approach. We specify prior distributions for the parameters Δ, Ω, and ξ and use Markov Chain Monte Carlo (MCMC) methods for obtaining draws from the posterior distribution of these parameters and functions thereof. The structure of the model is particularly well suited to such an approach. There are large numbers of parameters that can be treated as unobserved random variables and imputed in the MCMC algorithm. In addition, the likelihood function is likely to have multiple modes, implying that quadratic approximations to its shape are likely to be poor, resulting in poor properties of large sample confidence intervals for the underlying parameters. It should be noted though that these multiple modes need not make the normal approximation to the posterior distribution of the effects of policies of interest (e.g., price changes, or the market share of a new product) inaccurate. For example, one problem with frequentist inference in the current setting with at least two unobserved product characteristics is that these are never separately identified. This does not matter for most purposes because many estimands of interest would be invariant to the re-labelling of the unobserved product characteristics. However, if an asymptotic approximation is based on a quadratic approximation to the likelihood function in all its arguments, followed by the delta method, the results could be sensitive to such multiple modes. More generally, the numerical problems in locating the maximum or maxima of the likelihood function can be severe.

The implementation of the MCMC algorithm borrows heavily from Rossi, McCulloch and Allenby (1996), RMA hereafter, as well as more indirectly from work by Chib and Greenberg

(1998) on Gibbs sampling in latent index models. For a general discussion of MCMC methods see Tanner (1993), Gelman, Carlin, Stern and Rubin (2004), and Geweke (1997). Here we briefly discuss the general approach we take in this paper. An appendix posted on the web (Athey and Imbens, 2007) contains more details on the specific implementation.

The specific model we estimate is given in (5.7). Let Y_{it} denote the choice, $Y_{it} \in \{1, \ldots, J\}$. We observe T_i choices for individual i, each in a different market. For each of these choices we observe the product choosen, the product characteristics of the all the products in that market, X_{jm}, and the individual characteristics Z_{it}. We assume that conditional on ν_i, ξ_j, Z_{it}, and X_{jm} the idiosyncratic error term ϵ_{ijt} is normally distributed with mean zero and unit variance. Conditional on ξ_j, Z_{it}, and X_{jm} the unobserved individual component ν_i is normally distributed with mean zero and covariance matrix Ω.

In order to calculate the posterior distribution we need to specify prior distributions for common parameters Ω, Δ, and for the unobserved choice characteristics ξ_j. We use proper prior distributions for each parameter. The prior distribution on each element of Δ is normal with mean zero and variance 1/4. The elements of Δ are assumed to be independent a priori. The prior distribution on Ω is Wishart with parameters 100 and 0.01 times the $K + P$ dimensional identity matrix. The prior distribution on ξ_j is normal with mean zero and unit variance.

6 Application

6.1 Data

To illustrate the methods developed in this paper we analyze the demand for yogurt using scanner data from a market research firm (A.C. Nielsen) collected from 1985 through 1988. See Ackerberg (1998, 1999) for more information regarding these data. We focus on data from a single city, Springfield, Illinois. We restrict attention to purchases of a single-serving size. We excluded purchases where more than a single unit of yogurt was purchased.[6] Eight brands of yoghurt appear in the remaining dataset. We have a total of 16,824 purchases by 1038 households. These are divided over 21 stores during a period of 138 weeks. For each household we use a single observed household characteristic, household income. This is measured in 14 categories, ranging from 0-5,000 to more than 100,000. For each category we impute the mid-point of the category as the actual household income, with 125,000 for the highest (over 100,000) income category. Table 1 presents some summary statistics for this variable and for the number of purchases per household. We average the income over the 1038 households, weighted by the number of purchases per household.

For each yogurt brand we use two observed characteristics, price measured in cents[7] and

[6]We lose about one third of the observations due to this restriction. This is clearly a crude approach to dealing with the issues that arise in modeling multiple purchases, which may include multiple purchases of a single brand as well as purchases of more than one brand on a single trip. However, it simplifies the analysis and exposition of the application of the methods.

[7]We ignore the presence of coupons. Coupons are notoriously difficult to deal with because whether or not a consumer has access to a coupon is unobservable. It is possible to impute whether a coupon was in principle available in a market by checking whether any consumer used one for a particular product in a particular week, but not all consumers are aware of available coupons. See Osborne (2005) for a innovative way of estimating the

a binary indicator for whether the product was featured in advertising that week. For each purchase we directly observe these variables for the brand that was actually purchased. For our analysis we also need to know the values of these variables for the seven brands that were not purchased in that transaction for that particular market. We take the market to be a store in a particular week. We impute the price for the other seven brands by taking the average price for all purchases of each of these seven brands over all transactions for that brand in the same week and in the same store. We impute the feature variable as one if for any purchase of that brand in the same store in the same week the product was featured. Typically there was no recorded purchase for at least some of the eight brands during that week in that store. In that case we remove the brands for which there were no purchases from the choice set of the individual for that purchase. As a result the choice set varies in size across observations. On average there are 2.36 brands in a consumer's choice set on a trip in which the consumer purchased yogurt.

Table 2 reports summary statistics for the eight brands. We report averages over all purchases where the brand was included in the choice set, as well as over purchases of each brand. For example, the second row of Table 2 presents the information for the biggest brand, Dannon. Its market share is 49%. Its average price (averaged over all purchases where dannon was in the choice set) was 60.13 cents, rangeing from 20 cents to 73 cents. It was featured in the store during 9% of the purchases. It was in 88% of the choice sets. Averaged over all purchases of Dannon its price was 58.36 cents, slightly lower than the average over all purchases. It was more likely to be featured when it was purchased. On average there were 2.25 products in the choice set when Dannon was purchased.

6.2 Posterior Distribution of Parameters and Elasticities

We estimate four versions of the model. These versions are nested, so that it is straightforward to see the biases generated by placing unwarranted restrictions on the model. First we estimate the model with no unobserved product characteristics ($P = 0$), and with no unobserved individual characteristics ($\Omega = 0$). The second model allows for individual unobserved heterogeneity by freeing up Ω. The third model incorporates a single unobserved choice characteristic ($P = 1$). The fourth model allows for two unobserved product characteristics ($P = 2$).

In Table 3 we report the posterior distribution for selected parameters. First, we report the posterior mean and standard deviation for the average of the price coefficient β_{price}. We also report measures of the variation in this coefficient. We decompose this variation into the part due to variation in the observed individual coefficients and due to variation in the unobserved individual characteristics. We report the standard deviation of both components. We also report the summary statistics for the average and the two standard deviations of the feature coefficient β_{feature}. Finally, we report summary statistics of the posterior distribution of the effect of income on the price coefficient, $\Delta_{\text{price,income}}$, and the effect of income on the feature coefficient, $\Delta_{\text{feature,income}}$.

For the model with two unobserved product characteristics we see that on average, a higher

propensity to use coupons.

price lowers utility (the posterior mean of the average over all individuals of β_{price} is negative), but that there is considerable variation in the price coefficient between individuals. This variation is partly due to variation in the observed individual characteristics (a standard deviation of 0.233) and partly due to variation in the unobserved individual characteristics (a standard deviation of 0.463). On average being featured increases demand for a product. Invididuals with higher income are found to be less price-sensitive (the posterior mean of $\Delta_{\text{price,income}}$ is positive). With income measured in 10,000's of dollars, the point estimates suggest that individuals with a household income of \$60,000 have a price coefficient of approximately zero ($-4.09 + 60 \times 0.069 \approx 0$). (Recall from Table 1 that average household income in this data set is 35,000.) Income does not appear to have much of an effect on the relation between feature and demand.

It is interesting to note that with no unobserved choice characteristics the model estimates a much larger role for the feature variable. This would be consistent with the feature variable being a noisy measure for the unobserved product characteristics that actually matter for utility.

A potentially important difference between the estimates from model with two unobserved choice characteristics and the model with only one is that the estimated standard deviation of the price coefficient is larger in the model with one unobserved choice characteristic (.541 versus .463, with the standard deviations of these parameters equal to .021 and .027, respectively). This suggests that using a model that is too restrictive in terms of unobservable product characteristics can force estimates that imply too much heterogeneity in price sensitivity. For some counterfactuals, these differences might lead to inaccurate predictions. For example, using a model with only one unobserved product characteristic, the entry of a low-price, low-quality brand or a high-price, high-quality brand might lead to predictions of market shares for the new product that are too large.

Next we report own- and cross-price elasticities for the eight brands. To estimate the elasticities, we first estimate them for each individual conditional on the choice sets and the unobserved individual and choice characteristics. Then we average over all individuals. The results for the four models are in Tables 4–7. For the first two models we see large positive own price elasticities, as well as numerous (large) negative cross-price elasticities. For the model with one unobserved characteristic the elasticities have a few entries with unexpected signs and magnitudes. For the model with two unobserved product characteristics we see all of the own- and cross-price elasticities are of the expected sign, and they are of reasonable magnitudes (all own price elasiticies are larger than one in absolute value). For the largest brand, Dannon, the own-price elasticity is -5.37, and the cross-price elasticity of Dannon with respect to the second biggest is 2.84. These results suggest an important role for unobserved product characteristics, which could include the propensity to issue coupons (which were excluded from our model for simplicity), quality, flavor mix, and brand recognition.

6.3 Predicting Market Shares for New Products

To compare the counterfactual predictions arising from the different models, we simulate market shares for a new product. The product we introduce has the same observed characteristics in each market, a price equal to the average value of the price in the entire market (47 cts), and is never featured (feature= 0). It is included in every individual's choice set. For the first two models this information is sufficient to predict the market share. For the models with unobserved product characteristics we also need to specify values for the unobserved characteristics. As discussed in Section 4, we draw the unobserved choice characteristics randomly from the marginal distribution of unobserved choice characteristics estimated from the sample. This has the effect of making the predicted market shares more uncertain, so that even with an infinitely large sample we would not be able to predict the market share for the new product with certainty. Instead, there is a range of possible market shares, depending on the values of the unobserved characteristics.

The results for this exercise are in Table 8, where the additional variation from adding unobservable product characteristics is apparent. Perhaps surprisingly, there is little change in the estimates from including two versus one unobserved product characteristic.

6.4 Sensitivity to Choices for Prior Distributions

Here we investigate the sensitivity of the results to the specification of the prior distributions. We focus on the most general model with two unobserved choice characteristics that is most likely to be sensitive to this specification. For five different specifications of the prior distributions we report same parameter estimates as in Table 3, the own price for Dannon, and the cross-price elasticity for Dannon with respect to the price of CTL BR, and the summary statistics for the distribution of the predicted market share for a new product.

In the first pair of columns we report the results for the baseline prior distribution. Differences between these columsn and the previously reported results for the $P = 2$ model reflect on the lack of accuracy of the MCMC calculations (based on runs of 40,000 iterations). In the second pair of columns we change the prior variance for Δ from an identity matrix multiplied by 0.25 to an identity matrix multiplied by 0.125. In the third pair of columns we change the first parameter of the prior distribution of Σ from 100 to 50. In the fourth pair of columns we change the second parameter of the prior distribution of Σ from 0.01 to 0.1. The results are in Table 9. Generally the specification of the prior distributions changes the posterior distributions somewhat, but it does not change the qualitative conclusions.

7 Conclusion

This paper explores an issue first raised by McFadden (e.g., McFadden, 1981), namely the extent to which discrete choice models should incorporate unobserved product characteristics in order to rationalize choice data in settings with many products and/or multiple markets. We find that in general a model should include at least two unobserved choice characteristic if monotonicity of the utility function in the unobserved choice characteristics is imposed. More than two unobserved characteristics may be needed only if the functional form of the utility function (and in particular, its dependence on unobserved characteristics) is restricted.

We find that MCMC methods enable us to implement such models in a straightforward manner. We illustrate the method using scanner data about yogurt purchases. Our main findings are that the inclusion of two unobserved choice characteristics leads to more reasonable estimates of elasticities. We also argue that our approach leads to more realistic predictions about the heterogeneity in potential market shares that might arise on introduction of a new product. With additional structure, these predictions can be sharpened. In addition, the dependence of predicted market share on the location of a new product in characteristic space (both observable and unobservable characteristics) can be analyzed. We believe that an important advantage of the framework we propose is that the unobservable component of utility has a fair amount of structure, and the interpretability of the resulting estimates help guide the researcher in conducting counterfactual simulations. In applications, it may be possible to analyze and interpret the unobservable product characteristics, in order to gain a sense of how existing products are positioned and to help discover what parts of the product space might be most ripe for entry.

A number of questions are left open for future work. Among these is the question of how much individual heterogeneity is necessary to rationalize choice data in a variety of settings, and how that depends on any functional form or monotonicity restrictions that are imposed in the specification of individual utility.

References

Ackerberg, D., (1998), "Advertising, Learning and Consumer Choice in Experience Good Markets: A Structural Empirical Examination, "Department of Economics, UCLA.

Ackerberg, D., (1999), "Empirically Distinguishing Informative and Prestige Effects of Advertising, "Department of Economics, UCLA. .

Ackerberg, D., and M. Rysman, (2002), "Unobserved Product Differentiation in Discrete Choice Models: Estimating Price Elasticities and Welfare Effects, "Department of Economics, UCLA.

Ackerberg, D., L. Benkard, S. Berry, and A. Pakes, (2006), "Econometric Tools for Analyzing Market Outcomes,"forthcoming in J. Heckman and E. Leamer (eds), *Handbook of Econometrics*, Vol. 5B, Amsterdam, North Holland

Athey, S., and G.Imbens, (2007), "Appendix to: 'Discrete Choice Models with Unobserved Choice Characteristics',", Unpublished Manuscript paper 6600.

Allenby, G., Y. Chen, and S. Yang, (2003), "Bayesian Analysis of Simultaneous Demand and Supply," *Quantitative Marketing and Economics*, 1, 251-275.

Bajari, P., and L. Benkard, (2003), "Discrete Choice Models as Structural Models of Demand: Some Economic Implications of Common Approaches ", Stanford Business School.

Bajari, P., and L. Benkard, (2004), "Demand Estimation with Heterogenous Consumers and Unobserved Product Characteristics: A Hedonic Approach ", Stanford Business School.

Berry, S., (1994), "Estimating Discrete-Choice Models of Product Differentiation,"*RAND Journal of Economics*, Vol. 25, 242-262.

Berry, S., J. Levinsohn, and A. Pakes (1995), "Automobile Prices in Market Equilibrium, "*Econometrica*, 63(4), 841-889.

Berry, S., J. Levinsohn, and A. Pakes (2004), "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market, ", *Journal of Political Economy*, Vol 112(1), 68-105.

Berry, S., O. Linton, and A. Pakes, (2004), "Limit Theorems for Estimating the Parameters of Differentiated Product Demand Systems ", *Review of Economic Studies*, Vol. 71, 613-654.

Berry, S., and A. Pakes, (2002), "The Pure Characteristics Discrete Choice Model of Differentiated Products Demand ", unpublished manuscript, Dept of Economics, Yale University.

Bresnahan, T., M. Trajtenberg, and S. Stern, (1997), "Market Segmentation and the Sources of Rents from Innovation: Personal Computers in the Late 1980's,"*RAND Journal of Economics*, Vol. 28(0), S17-S44.

Chib, S., (2003), "Markov Chain Monte Carlo Methods: Computation and Inference,"in Heckman and Leamer (eds), *Handbook of Econometrics*, Vol 5, Amsterdam, North Holland

Chib, S., and E. Greenberg (1998), "Analysis of Multivariate Probit Models,"*Biometrika*.

Elrod, T., and M. Keane, (1995), "A Factor-Analytic Probit Model for Representing the Market Structure in Panel Data "*Journal of Marketing Research*, Vol. XXXII, 1-16.

Gelman, A., J. Carlin, H. Stern, and D. Rubin, (2004), *Bayesian Data Analysis* Chapman and Hall.

Geweke, J., (1997), "Posterior simulators in econometrics,"in Kreps and Wallis (eds) *Advances in Economics and Econometrics: Theory and Applications*, Cambridge University Press, Cambridge, UK.

Geweke, J., and M. Keane (2002), "Bayesian Inference for Dynamic Discrete Choice Models without the Need for Dynamic Programming,"in *Simulation Based Inference and Econometrics: Methods and Applications*, Mariano, Schuermann and Weeks (eds.), Cambridge University Press, 100-131.

Goettler, J., and R. Shachar (2001), "Spatial Competition in the Network Television Industry,"*RAND Journal of Economics*, Vol. 32(4), 624-656.

Goldberg, P., (1995), "Product Differentiation and Oligopoly in International Markets: The Case of the U.S. Automobile Industry ", *Econometrica*, 63(4), 891-951.

Harris, K., and M. Keane, (1999), "A Model of Health Plan Choice: Inferring Preferences and Perceptions from a Combination of Revealed Preference and Attitudinal Data,"*Journal of Econometrics*, 89, 131-157.

Hausman, J., and D. McFadden, (1984), "Specification Tests for the Multinomial Logit Model ", *Econometrica* 52(5), 1219-1240.

Keane, M. (1997), "Modeling Heterogeneity and State Dependence in Consumer Choice Behavior,"*Journal of Business and Economic Statistics*, 15(3), 310-327.

Keane, M. (2004), "Modeling Health Insurance Choice Using the Heterogenous Logit Model,"unpublished manuscript, department of economics, Yale University.

Kim, J., G. Allenby, and P. Rossi, (2006) "Product Attributes and Models of Multiple Discreteness,"*Journal of Econometrics* forthcoming.

Manski, C. (2003), *Partial Identification of Probability Distributions*, New York: Springer-Verlag.

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

McCulloch, R., N. Polson, and P. Rossi, (2000) "A Bayesian Analysis of the Multinomial Probit Model with Fully Identified Parameters, "*Journal of Econometrics* 99, 173-193.

McFadden, D., (1973), "Conditional Logit Analysis of Qualitative Choice Behavior "in P. Zarembka (ed), *Frontiers in Econometrics* Academic Press, New York 105-142.

McFadden, D., (1981), "Econometric Models of Probabilistic Choice, "in Manski and McFadden (eds), *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press, Cambridge, MA.

McFadden, D., (1982), "Qualitative Response Models "in Hildenbrand (eds), *Advances in Econometrics: Invited Papers for the Fourth World Congress of the Econometric Society*, 1-37, Cambridge University Press, Cambridge, UK.

McFadden, D., (1984), "Econometric Analysis of Qualitative Response Models,"in Griliches and Intriligator (eds), *Handbook of Econometrics*, Vol. 2, 1395- 1457, Amsterdam, North Holland

McFadden, D., (1989), "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration, "*Econometrica* 57(5) 995-1026. .

Nevo, A., (2000), "A Practitioner's Guide to Estimation of Random-Coefficients Logit Models of Demand,"*Journal of Economics and Management Strategy* 9(4) 513-548.

Nevo, A., (2001), "Measuring Market Power int he Ready-to-eat Cereal Industry,"*Econometrica* 69(2) 307-342.

Osborne, M., (2005), "Consumer Learning, Habit Formation and Heterogeneity: A Structural Examination," Mimeo, Stanford University.

Pakes, A., and D. Pollard, (1989), "Simulation and the Asymptotics of Optimization Estimators,"*Econometrica* 57(5) 1027-1057 .

Petrin, A., (2002), "Quantifying the Benefits of New Products: The Case of the Minivan "*Journal of Political Economy*, Vol. 110(4), 705-729.

Petrin, A., and K. Train (2005), "Tests for Omitted Attributes in Differentiated Product Models,"Unpublished Manuscript, University of Chicago Business School.

Poole, K., and H. Rosenthal, (1985), "A Spatial Model for Legislative Roll Call Analysis, "*American Journal of Political Science*, Vol. 29, 357-384.

"Organization, Econometrics, Vol 6,

Romeo, C., (2003), "A Gibbs Sampler for Mixed Logit Analysis of Differentiated Product Markets Using Aggregate Data "unpublished manuscript, Economic Analysis Group, US Department of Justice.

Rossi, P., R. McCulloch, and G. Allenby, (1996), "The Value of Purchase History Data in Target Marketing," Marketing Science, 15(4), 321-340.

Rossi, P., G. Allenby, and R. McCulloch (2005), *Bayesian Statistics and Marketing*, John Wiley and Sons.

Tanner M., (1993), *Tools for Statistical Inference : Methods for the Exploration of Posterior Distributions and Likelihood Functions* Springer.

Tanner M., and W. Wong, (1987), "The Calculation of Posterior Distributions by Data Augmentation,"*Journal of the American Statistical Association*, Vol 82(398) 528-540.

Train, K., (2003), *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge U.K.

Table 1: Summary Statistics: Individual Characteristics

Characteristic	mean	stand dev	minimum	maximum
number of purchases	16.21	24.04	1.00	285.00
household income	34.47	22.98	2.50	125.00

The first row gives summary statistics for the number of purchases for the 1038 households. The second row gives summary statistics for income per household, weighted by the number of purchases per household. Total number of purchases is 16,824.

Table 2: Summary Statistics: Choice Characteristics

	averaged over all transactions							brand purchases		
	market	price				feature	incl in	price	feature	ave size
Brand	share	ave	st dev	min	max	ave	choice set	ave	ave	choice set
Wght Wtch	0.09	62.74	6.79	25.00	73.00	0.04	0.41	61.33	0.09	2.73
Dannon	0.49	60.13	9.84	20.00	73.00	0.09	0.88	58.36	0.13	2.25
Elmgrove	0.04	30.94	3.56	22.00	33.00	0.15	0.11	29.84	0.25	2.77
YAMI	0.04	32.06	9.54	20.00	59.00	0.21	0.11	29.18	0.36	2.56
HWT MDY	0.06	30.72	3.54	25.00	33.00	0.13	0.15	29.33	0.22	2.16
HILAND	0.10	37.20	9.62	20.00	55.00	0.16	0.24	35.49	0.21	2.47
NTRL LE	0.02	32.40	10.39	20.00	55.00	0.11	0.11	32.81	0.09	3.10
CTL BR	0.17	35.65	5.79	20.00	45.00	0.20	0.36	34.47	0.23	2.30

Column 2 reports the market share of the brand in this data set. Columns 3-6 report the average price over all store/weeks in which this brand was in the choice set, as well as the standard deviation, minimum and maximum. Column 7 reports the fraction of the times the brand was featured. Column 8 reports the fraction of the store/week combinations that the brand was in the choice set (had at least one purchase in that market) Columns 7 and 8 report averages for price and feature variable over the all purchases of the brand. Column 9 gives the average size of the choice set during the purchases of that brand.

Table 3: Summary Statistics Posterior Distribution for Selected Parameters

Parameter	$P=0, \Omega=0$ mean	std	$P=0$ mean	std	$P=1$ mean	std	$P=2$ mean	std
mean(β_{price})	-0.012	0.004	-0.007	0.007	-0.302	0.020	-0.409	0.020
std ($\Delta_{\text{price,income}} \cdot Z_{\text{income}}$)	0.253	0.005	0.289	0.041	0.230	0.039	0.233	0.039
$\sqrt{\Omega_{price,price}}$	0	0	0.586	0.022	0.541	0.021	0.463	0.027
mean(β_{feature})	0.663	0.026	0.743	0.041	0.449	0.051	0.379	0.047
std ($\Delta_{\text{feature,income}} \cdot Z_{\text{income}}$)	0.111	0.008	0.379	0.100	0.070	0.034	0.067	0.037
$\sqrt{\Omega_{feature,feature}}$	0	0	0.983	0.070	0.133	0.017	0.153	0.026
$\Delta_{\text{price,income}}$	0.048	0.003	0.060	0.010	0.053	0.011	0.069	0.010
$\Delta_{\text{feature,income}}$	-0.007	0.013	-0.025	0.032	-0.028	0.017	-0.011	0.024

Column 2-3 report the mean and standard deviation for various parameters for the model with no unobserved choice characteristics ($P = 0$) and no unobserved individual heterogeneity ($\Omega = 0$). Column 4-5 report the mean and standard deviation for the same parameters for the model with no unobserved choice characteristics ($P = 0$) allowing for unobserved individual heterogeneity ($\Omega \neq 0$). Column 6-7 report the mean and standard deviation for the same parameters for the model with a single unobserved choice characteristics ($P = 1$) allowing for unobserved individual heterogeneity ($\Omega \neq 0$). Column 6-7 report the mean and standard deviation for the same parameters for the model with two unobserved choice characteristics ($P = 2$) allowing for unobserved individual heterogeneity. The parameters reported on include the average effect of price on the utility, the standard deviation of the component of that effect corresponding to the observed individual characteristics and the standard deviation of the component of that effect corresponding to the unobserved individual characteristics, the same three parameters for the feature variable, and the effect of the interactions of income and price and income and feature on utility. The price is measured in dollars.

Table 4: Elastiticies for Model with No Unobserved Product Characteristics and No Unobserved Individual Heterogeneity ($P = 0,\ \Omega = 0$)

With Respect to →	Wght W	Dannon	Elmgr	YAMI	HWT	HILA	NTRL	CTL
Wght Wtch	4.94	-5.99	0.93	-0.16	1.18	0.18	-0.04	0.36
Dannon	-5.74	1.22	0.84	0.50	0.73	-0.06	-0.03	0.30
Elmgrove	1.10	1.45	-1.83	0.00	2.67	2.65	2.10	2.10
YAMI	-0.23	0.91	0.00	-1.19	1.86	-0.29	-0.65	1.13
HWT MDY	1.31	1.29	2.69	2.91	-1.29	0.00	0.00	4.37
HILAND	0.28	-0.10	1.98	-0.25	0.00	-1.23	2.46	1.68
NTRL LE	-0.07	-0.05	1.68	-0.63	0.00	3.29	-2.84	1.46
CTL BR	0.49	0.42	2.45	1.24	5.12	1.57	1.34	-0.51

Each row reports average elasticities for one product with respect to its own price and with respect to the price of the seven other products. These elasticities are calculated at the individual level for all markets that had both products in the choice set and then averaged over all those markets weighted by the number of transactions per market. A "–" indicates that there were no markets (store/week combinations) with both products.

Table 5: Elastiticies for Model with No Unobserved Product Characteristics $P = 0,\ \Omega \neq 0$)

With Respect to →	Wght W	Dannon	Elmgr	YAMI	HWT	HILA	NTRL	CTL
Wght Wtch	16.16	-17.93	0.72	-1.32	1.06	0.20	-0.21	0.28
Dannon	-16.12	4.82	0.92	0.28	0.27	-0.51	-0.32	0.10
Elmgrove	0.74	1.70	-3.13	0.00	9.53	5.96	5.34	5.14
YAMI	-2.04	0.52	0.00	-2.20	4.82	4.56	-1.89	2.99
HWT MDY	1.08	0.50	9.60	8.89	-1.94	0.00	0.00	6.92
HILAND	0.31	-0.89	4.71	2.51	0.00	-3.17	7.56	4.60
NTRL LE	-0.34	-0.72	4.28	-1.81	0.00	9.60	-7.97	4.46
CTL BR	0.38	0.14	6.63	3.15	12.71	3.80	3.57	-0.61

Table 6: ELASTITICIES FOR MODEL WITH A SINGLE UNOBSERVED PRODUCT CHARACTERISTIC $P = 1,\ \Omega \neq 0$)

With Respect to →	Wght W	Dannon	Elmgr	YAMI	HWT	HILA	NTRL	CTL
Wght Wtch	1.52	-3.19	0.96	0.34	2.10	0.45	0.11	1.02
Dannon	-1.23	-4.28	2.60	3.66	2.66	3.03	1.15	3.64
Elmgrove	0.68	6.40	-6.09	0.00	9.86	10.78	6.60	7.51
YAMI	0.23	9.99	0.00	-7.32	5.68	10.92	1.90	4.57
HWT MDY	1.37	7.23	9.64	13.93	-5.22	0.00	0.00	8.97
HILAND	0.52	6.68	6.79	3.67	0.00	-7.54	8.77	7.06
NTRL LE	0.13	3.75	5.41	2.18	0.00	13.56	-13.12	6.26
CTL BR	0.84	7.34	12.01	4.78	16.78	5.26	3.72	-4.19

Table 7: ELASTITICIES FOR MODEL WITH TWO UNOBSERVED PRODUCT CHARACTERISTICS $P = 2,\ \Omega \neq 0$)

With Respect to →	Wght W	Dannon	Elmgr	YAMI	HWT	HILA	NTRL	CTL
Wght Wtch	-6.40	2.92	1.83	3.79	3.02	2.04	1.83	2.84
Dannon	1.17	-5.37	2.34	3.94	2.85	3.60	1.60	3.54
Elmgrove	1.48	5.73	-5.60	0.00	8.78	8.82	4.48	7.42
YAMI	2.21	11.15	0.00	-7.72	5.10	9.12	3.56	3.57
HWT MDY	1.90	7.34	10.16	10.32	-5.10	0.00	0.00	9.34
HILAND	2.58	8.21	7.04	3.81	0.00	-8.49	8.68	8.03
NTRL LE	2.70	5.75	5.01	3.17	0.00	14.39	-14.78	6.29
CTL BR	2.44	7.17	12.63	3.60	21.83	5.59	3.15	-4.51

Table 8: PREDICTED MARKET SHARE FOR NEW PRODUCT

	$P = 0, \Omega = 0$	$P = 0$	$P = 1$	$P = 2$
average	0.254	0.201	0.241	0.254
standard deviation	0.001	0.001	0.110	0.125
0.05 quantile	0.252	0.199	0.117	0.116
0.95 quantile	0.256	0.203	0.360	0.394

The first row contains posterior means for the market share for a new product that is available in each market (each store/week), always with a price of 47 cents and not featured. For the models with unobserved product characteristics we draw the unobserved product characteristic(s) from their estimated marginal distribution. The second row gives the posterior standard deviation of this market share, and the third and fourth rows give the 0.05 and 0.95 quantiles of this posterior distribution.

Table 9: Sensitivity of Posterior Distributions to Prior Distribution

Parameter	Base Prior		Prior II		Prior III		Prior IV		Prior V	
mean(β_{price})	-0.40	0.02	-0.39	0.02	-0.42	0.02	-0.41	0.02	-0.40	0.02
std ($\Delta_{\text{price,income}} \cdot Z_{\text{income}}$)	0.26	0.03	0.25	0.04	0.22	0.08	0.26	0.07	0.26	0.04
$\sqrt{\Omega_{price,price}}$	0.49	0.04	0.46	0.03	0.62	0.05	0.56	0.03	0.48	0.03
mean(β_{feature})	0.40	0.05	0.405	0.05	0.37	0.07	0.38	0.06	0.39	0.05
std ($\Delta_{\text{feature,income}} \cdot Z_{\text{income}}$)	0.07	0.04	0.08	0.04	0.19	0.14	0.23	0.09	0.08	0.04
$\sqrt{\Omega_{feature,feature}}$	0.15	0.03	0.16	0.03	1.00	0.12	0.61	0.08	0.17	0.03
$\Delta_{\text{price,income}}$	0.07	0.01	0.06	0.01	0.08	0.02	0.07	0.02	0.06	0.01
$\Delta_{\text{feature,income}}$	-0.01	0.03	-0.01	0.03	-0.01	0.04	-0.01	0.03	-0.01	0.03
cross elast D wrt CRT BL	3.66	0.27	3.49	0.27	3.61	0.25	3.72	0.27	3.62	0.30
own elast D	-5.36	0.35	-5.27	0.30	-4.88	0.33	-5.12	0.24	-5.31	0.28
Market share new product	0.25	0.13	0.26	0.12	0.26	0.12	0.25	0.12	0.25	0.12

.