Demand Estimation With Heterogenous Consumers and Unobserved Product Characteristics: A Hedonic Approach.

Pat Bajari and Lanier Benkard
Stanford University
Preliminary: May contain errors. Comments encouraged.

April 11, 2001

Abstract

In this paper we study the problem of identification and estimation of preferences in pure hedonic models of demand for differentiated products. We argue that many commonly used random utility models impose a priori undesirable properties for welfare and substitution effects as the number of products becomes large, while the pure hedonic model does not. We study the integrability problem in the pure hedonic model with an unobserved product characteristic. We demonstrate that, unlike the case where all product characteristics are observed, it is not in general possible to uniquely recover consumer preferences from knowledge of the demand function. We state two special cases, however, when preferences can be recovered uniquely. We also consider the problem of identification and estimation when the entire demand function is not observed. If the product space is continuous, we propose a variant of the two-stage hedonic approach of Rosen (1974) to estimate preferences. If the product space is discrete, we propose a Gibbs sampling algorithm to simulate the posterior distribution of random coefficients.

1 Introduction

In this paper we study the problem of identification and estimation of preferences in pure hedonic models of demand for differentiated products. The paper's primary goal is recovery of the distribution of preferences in a population using standard data sets on prices and quantities and the characteristics of products in a narrowly defined market. Recovery of the distribution of preferences is important for two reasons. The first is that knowledge of the distribution of preferences allows researchers to analyze the distribution of welfare effects from a policy change. For example, we may be interested in learning the distributional impact of technological change, or the distributional impact of price changes. The second is that if the distribution of preferences is estimated with few restrictions, then it may be possible to more accurately estimate the aggregate demand function (using explicit aggregation) than it would be using standard approaches.

Many recent papers in the empirical industrial organization literature (e.g., Ackerberg and Rysman (2000), Berry (1994), Berry and Pakes (2000), Berry, Levinsohn and Pakes (1995, 1998) [BLP, BLP2], Davis (2000), Hendel (1999), Nevo (2000), and Petrin (1998)) have attempted to find better ways to estimate demand systems in markets with differentiated products. This literature has generalized benchmark models such as the logit and nested logit in two primary ways. First, in order to make the demand systems more flexible and to avoid restrictive IIA assumptions on aggregate demand, these papers have estimated demand systems with random coefficients. Second, recent work has focused on strategies for identifying a product fixed effect in random utility models. Many authors (e.g. Berry (1994), BLP) have demonstrated that if unobserved product characteristics are positively correlated with price, estimating a demand system that ignores this correlation, such as the standard logit or nested logit, results in downwardly biased estimates of price elasticities. This bias inevitably leads to incorrect measures of welfare effects, substitution effects, and market power.

These papers have made a very important contribution to demand estimation, and we hope to preserve in this paper the lessons learned by this literature. However, in our opinion, along with that of many other authors (e.g. Berry and Pakes (2000), Ackerberg and Rysman (2000)), there are still at least four limitations in the current technology with respect to the goals of this paper.

First, in order to obtain consistent estimates of the demand system it is necessary to find instruments that are correlated with the endogenous variables, such as price, but uncorrelated with the product attribute that is unobserved to the econometrician. Thus far, it has proven difficult to find compelling instruments. Many papers (BLP, BLP2, Davis (2000), Petrin (1998) and others) use observed product characteristics as instruments. However, these papers admit that the assumption that unobserved product characteristics are mean independent of observed characteristics is not very compelling. Nevo (2000) improves on these instruments by using prices of goods in other markets and market dummies as instruments. However, Nevo (2000) also notes that there are reasonable sets of assumptions under which each of these alternative sets of instruments would be invalid.

Second, it is unclear what role parametric assumptions, e.g., in the utility function and the distribution of random coefficients, play in identifying the model. Typically, papers in this literature estimate a model where the consumer's utility function is linear in the product attributes, including an additive logit error term, and the random coefficients are normally distributed. What is unclear is whether these models are identified without these parametric assumptions. Furthermore, if our goal is to recover the distribution of preferences, such restrictions may be undesirable.

Third, many random utility models such as the logit and nested logit have been found to perform poorly in practice, particularly when there are a large number of products in the market. In applications such as BLP and Petrin (1998), the logit model was found to imply welfare effects and substitution patterns that were a priori implausible.

Lastly, random utility models with unobserved product characteristics are computationally burdensome to estimate, particularly when the number of choices becomes large. Indeed, it is not clear that it would even be possible, using current computing technology, to estimate a BLP model for markets such as computers or housing where there are thousands of choices in the choice set.[1] The computational difficulty of estimating these models also discourages some researchers from applying these methods even though they are sympathetic to the approach.

In this paper, we study the identification and estimation of *pure hedonic* models of demand. In the pure hedonic model, unlike random utility models, consumer demand is deterministic. Each product is characterized by a finite number of attributes and the consumer is a rational utility maximizer. What distinguishes our work from much of the previous literature on pure hedonic models (one exception is Berry and Pakes (2000)) is that we focus on the case where the economist does not observe all of the product characteristics. We argue that in some applications the pure hedonic model may be preferable to random utility models. We also attempt to make some headway on the four limitations mentioned above in part by using a pure hedonic model instead of a random utility model.

This paper also makes several contributions to the literature on demand estimation with differentiated products. In section 2, we study the behavior of a broad class of random utility models as the number of products becomes large. We prove a set of theorems that demonstrate that these models have counterintuitive implications for welfare and substitution patterns as the number of products in the market becomes large. We consider these theorems to be most relevant to applications evaluating counterfactuals in which the number of products in the market changes. The conclusion of this section is that standard models may give misleading results in such cases.

In section 3, we study the problem of identification of preferences in the pure hedonic model when one product characteristic is not observed by the economist. We establish that, in general, even if a consumer's entire demand function is known, it will not be possible to uniquely recover consumer preferences. In fact, it is possible in a very wide class of models to attribute all of the consumer's utility to the unobserved product characteristics. This result, which is similar to that of Varian (1988), stands in sharp contrast to the integrability results of Hurwicz and Uzawa (1971) which suggest that preferences can be recovered from observed choices.

We go on to establish that the consumer's preferences can be recovered in at least two special cases. The first case we study is when unobserved product characteristics are independent of the observed product characteristics. This is similar to the case that the previous literature on differentiated products demand estimation has concentrated on, but slightly more restrictive.

[1] Bajari and Kahn (2000) use a BLP style demand model for the housing market, overcoming the high dimension of the choice set by grouping houses into a much smaller number of housing classes.

Using the results of Matzkin (1999), we demonstrate that in this case, under some weak conditions, it is possible to uniquely recover a consumer's preferences. Unlike the previous literature, we allow for completely general functional forms for both the utility function and the equilibrium pricing function. We also generalize the model to allow for an additive "measurement error" in prices, because we believe that typically prices are measured poorly in the types of applications we are interested in. We show that even if prices are measured with error, the price function and the consumer's preferences are identified.

Because we believe the independence assumption to be quite strong, we also consider a second case, in which we think of the consumer's maximization problem as one where she first chooses a "model" and second chooses a set of "options packages". Many product markets have this feature, such as automobiles or computers. For example, the models in the auto market include the Camry, Jetta, and Taurus and the options packages include horsepower, power steering, air conditioning and so forth. If the unobserved product characteristic is isomorphic to the "model" then we demonstrate that it is possible to identify the unobserved product characteristic (up to a monotone transformation) and uniquely recover the consumer's preferences if the entire demand function is known. Once again, we allow for completely general functional forms for both the equilibrium pricing function and the utility function, and we show that identification holds even if prices are measured with error.

In section 3.7, we consider the identification problem when the entire demand function is not observed. While it is useful to study the case where the entire demand function for a consumer is observed, this is the exception rather than the rule for applied studies. More typically, the econometrician may have only a handful or even a single observation per consumer. For example, this would typically be the case in applications using aggregate data.[2] In such cases, if the economist can consistently estimate the hedonic price surface that relates product characteristics to prices, then for each choice observed for a given individual it is possible to recover the consumer's marginal rate of substitution at the chosen bundle. We derive conditions under which some simple structural assumptions, such as assumptions on the functional form of the utility function, allow the econometrician to infer an individual's entire utility function from knowledge of the marginal rates of substitution at a finite number of chosen bundles. In many of the random coefficient models that have been used in the recent literature (e.g. BLP, BLP2, Petrin (1998), Nevo (2000)), knowledge of the marginal rates of substitution at a single chosen bundle is enough to recover each individual's taste coefficients, and thus the entire utility function.

Throughout this paper, identification of preferences is obtained through the existence of an equilibrium price surface. In section 3.2 we show that if demand is given by the pure hedonic model, then prices in each market can be written as a function of the observable and unobservable characteristics of the products in that market.[3] We also show that under some

[2]In fact there may be many observations per individual in aggregate data, but because there is typically no way to link these observations, researchers often assume that each unit sold in aggregate data corresponds to a different individual with independent preferences.

[3]Note that the price surface is simply a mapping from characteristics to a unique price, *conditional on the primatives of the market*. The price surface depends on the number of products in the market, their

weak conditions the equilibrium price surface must be Lipschitz continuous in characteristics, and strictly increasing in the unobserved characteristic. The price surface is used to identify and estimate the values of the unobserved product characteristics. Once the values of the unobserved product characteristics are obtained, identification of preferences becomes equivalent to the standard integrability problem. Since our arguments showing the existence of an equilibrium price surface are demand based and must be satisfied by any supply-side equilibrium, we believe our results to be quite general, extending to both static and dynamic contexts.

In sections 4.1-4.2, we propose a two-stage estimation procedure in the spirit of the two-stage procedure of Rosen (1974), where we first estimate the relationship between prices and product characteristics (both observed to the economist and unobserved) and second recover the consumer's random coefficients from observed choices. This approach allows the economist to estimate the distribution of random coefficients and the unobserved product characteristics, in some cases with weaker assumptions than the previous literature. Our proposed estimation procedure has the advantage of being fairly general and having a low computational burden. The cost of this generality is that our procedure is more demanding of the data than the techniques used in the previous literature, and thus is not likely to work well for markets with small numbers of products.

Lastly, in section 4.3, we turn to the important case where the product space is discrete instead of continuous. In this case, an individual consumer's taste coefficients will not typically be identified even if the entire demand function is observed. Instead, each individual's taste coefficients can be shown to lie in a set. We also show that in general this set is smaller when there are more products in the market, and that the set converges to the individual's taste coefficients when the number of products becomes large. To estimate the aggregate distribution of preferences, we develop a simple Gibbs sampling procedure. The Gibbs procedure is shown to converge to the population distribution of taste coefficients when characteristics are continuous and the number of products becomes large. The Gibbs procedure also has a low computational burden.

Our work builds on several literatures in microeconomics and applied microeconomics. The first deals with estimating discrete choice models with random coefficients and unobserved product characteristics. Recent papers that estimate such models include Berry and Pakes (2000), Berry, Levinsohn and Pakes (1995, 1998), Davis (2000), Goettler and Shachar(1999), Hendel (1999), Nevo (2000) and Petrin (1998). The second is the literature that uses a two step hedonic procedure to estimate preferences for differentiated products. This includes Rosen (1974), Epple (1987) and Bartik (1987). The third is the literature on revealed preference and integrability, for instance Richter (1966), Hurwicz and Uzawa (1971), and Varian (1988). Lastly, we rely on recent work on nonparametric estimation of econometric models without additively separable error terms, including Blundell and Powell (2000), and particularly Matzkin (1999).

characteristics, which firms are producing them, etc., and the price surface would shift if any of these primitives were changed. See section 3.2 for further discussion.

2 Economic Implications of the Standard Discrete Choice Econometric Models

This section outlines several properties of the standard discrete choice econometric models (GEV, Logit, Probit, etc.) that seem to us to be undesirable from the point of view of economic theory and/or basic economic intuition. We include this section in this paper because it motivates our decision to use an alternative model to those used in the previous literature. We are by no means the first to raise many of these issues (see e.g. BLP, BLP2, Berry and Pakes (2000), Andersen De-Palma and Thisse (1992), Petrin (1998), Caplin and Nalebuff (1991)). However, we believe that some of the properties listed here are new to the literature, and furthermore we have attempted to show all of our results formally.

We begin by introducing some notation and by making some simplifying assumptions. Note that the assumptions made in this section are designed to relate to the existing literature in I.O. on demand systems. We will not carry these assumptions forward to later sections.

Following the previous literature, we assume that individuals' utility functions over products can be written as a function of individual characteristics (describing individual tastes), product characteristics, and an additively separable random error term:

$$u_{ij} = u_0(y_i) + u(x_j, y_i - p_j, \beta_i) + \epsilon_{ij} \qquad \text{for } j \in 1..J \tag{1}$$

where x_j is a vector of product characteristics, p_j is product price, y_i is income, β_i is a vector of individual taste parameters, and ϵ_{ij} is an individual and product specific random error. This utility function should be thought of as a direct utility function with preferences over the characteristics of inside goods and some composite outside good, with the budget constraint substituted in. Note that the $u_0(y_i)$ term has no impact on demand. We assume that the researcher does not observe the income of each individual, but knows its distribution F_y in the population. We assume that the taste coefficients (β_i) are random with some distribution $F_\beta(\cdot;\theta)$, where θ is a parameter vector. We also assume that there is an outside good with utility given by:

$$u_{i0} = u_0(y_i) + \epsilon_{i0} \tag{2}$$

It is not necessary to have an outside good in the model for most of what follows. We include it because its presence underscores some of the undesirable properties of the model, and because most of the models in the previous literature contain an outside good with similar specification.

The random error term is typically introduced as an econometric tool. Under some common assumptions it guarantees that at every parameter vector all choices have positive probability for every individual. Therefore, the model has the ability to rationalize any data set, and all values of the parameters have positive likelihood. The error term is typically explained as representing some unobservable tastes and/or product characteristics. We investigate the validity of this assertion in more detail below.

We now introduce an assumption on the error term that is common in the previous literature. We denote the vector of error terms for individual i excluding the error term for product j as $\epsilon_{i,-j}$.

Assumption R For all $M < \infty$, there exists a δ_M such that $Pr(\epsilon_{ij} < M|\epsilon_{i,-j}) < \delta_M < 1$ for all i, j, all $\epsilon_{i,-j}$, and all $J \in \mathcal{Z}_+$.

Assumption R ensures that the conditional error distributions are neither collapsing to a point nor diverging to minus infinity as more products are added to the market.[4] It guarantees that $\lim_{J\to\infty} \max_{j\in 1..J} \epsilon_{ij} = \infty$ $a.s.$[5]. Assumption R is maintained in all applications of discrete choice that we are aware of. In particular, all GEV models, the probit model, and all GEV- and Probit-based random coefficients models satisfy R, so long as the errors have positive variance and are not perfectly correlated.[6]

We also need some conditions on the set of utility functions that we will consider. Let $X \subseteq \mathbb{R}^K$ be the support of x, and $B \subseteq \mathbb{R}^B$ be the support of β. The following assumptions are intended to match those made in the previous literature.

Assumption Ua

(i) $u(x, c, \beta)$ is continuous in all its arguments, and for all $(x, \beta) \in X \times B$, $u(x, \cdot, \beta)$ is strictly increasing.

(ii) $u_0(\cdot)$ is continuous and strictly increasing.

(iii) For all $(x, \beta) \in X \times B$, $lim_{c\to 0^+} u(x, c, \beta) = -\infty$. For all $c < 0$ and all $(x, \beta) \in X \times B$, $u(x, c, \beta) = -\infty$.

(iv) F_y has full support on $\mathbb{R}^+$.

Assumption Ub

(i) $u(x, c, \beta)$ is continuous in all its arguments, and for all $(x, \beta) \in X \times B$, $u(x, \cdot, \beta)$ is strictly increasing.

(ii) $u_0(\cdot)$ is continuous and strictly increasing.

(iii) For all $(x, \beta) \in X \times B$, $lim_{c\to -\infty} u(x, c, \beta) = -\infty$.

[4] Assumption R is similar to requiring continuity in the upper tail of the conditional distributions. However, it does not necessarily require continuity in the upper tails since it would also be satisfied by any distribution with positive mass at infinity.

[5] See appendix for proof.

[6] Assumption R also requires that the error terms have means that are bounded away from minus infinity.

Assumption Ua is satisfied by BLP and Petrin(1998). Assumption Ub is satisfied by BLP2, Davis (2000), and Nevo (2000). The only difference between the two is whether or not there are income effects in demand for the inside good.

Together, these assumptions necessarily imply many economic properties which may not be desireable in empirical applications. The first thing to note about this model is that, for every product, demand is positive for every price.

1. **Demand is Positive for Every Price** *If assumption R holds, and either Ua or Ub holds, then in the model described above, the "virtual price", the price at which demand for the product equals zero, is equal to infinity for every product. That is, for every product, demand is positive for every price.*

This first property shows that standard discrete choice econometric models imply a very particular, and perhaps undesirable, shape for product level demand curves. The model enforces some differentiaton across all products, regardless of the similarity of their characteristics. There are several potential consequences of this differentiation. First, because product-level demand curves never touch the vertical (price) axis, and consumer surplus equals the area underneath the demand curve, it seems likely that the model is biased toward predicting large amounts of consumer surplus from each product. In fact, all of the welfare studies that we know of using models of this form find large amounts of welfare from each variety of the good (see, Petrin (1998), Trajtenburg (1989) and others). If we are trying to evaluate the effects of a policy change, we may not want to place such tight restrictions on the shape of the demand curve. Additionally, in the case of Ub, the implied demand system is not consistent with consumer theory, since it implies that some consumers must violate their budget constraints.[7]

Property one also implies that there does not necessarily exist a price surface in this model. That is, it is possible to have two different firms sell identical products (products with the same characteristics) at different prices in the same market.[8] By property one, both products would have positive demand.

We now list one additional assumption regarding the error distribution:

Assumption H Let $F_{\epsilon_j}(\cdot)$ denote the distribution of ϵ_{ij} conditional on $\epsilon_{i,-j}$. Assume that the limit as ϵ tends to the upper limit of its support of the hazard rate of $F_{\epsilon_j}(\cdot)$ is infinite, i.e., that $\lim_{\epsilon \to b} \frac{f_{\epsilon_j}(\epsilon)}{1-F_{\epsilon_j}(\epsilon)} = \infty$, where b is the upper end of the support of $F_{\epsilon}(\cdot)$ and b may equal infinity.

[7]This statement implicitly assumes that all consumers have strictly finite income and wealth.

[8]For example, this would happen if the two firms had different costs, or if the two firms were multi-product firms with different product lines and one identical product.

Assumption H is satisfied by all bounded distributions and the normal distribution (probit), but not the logit. It turns out that whether or not assumption H holds determines some key theoretical properties of the choice model.

We will now show that the behavior of existing models is particularly counterintuitive when the number of products in the market changes. To emphasize this, we concentrate on the properties of the econometric models in the extreme case when the number of products in the market tends to infinity. We have intentionally omitted the process by which the products are added when the limit is taken because the properties listed hold regardless of what process is generating the added products so long as the assumptions above are satisfied.

2. **Share of the Outside Good** *Let s_0 denote the market share of the outside good. If R holds, and either Ua or Ub holds, then either $s_0 \to 0$, or $s_0 > 0$ and $u(x_j, y_i - p_j, \beta_i) \to -\infty$ for all but a finite number of products for each individual that purchases the outside good.*
3. **Lack of Perfect Substitutes** *If R holds, and either Ua or Ub holds, and H does not hold, then in the limit each product almost surely does not have a perfect substitute. That is, each individual would suffer utility losses that are almost surely bounded away from zero if her first choice product were removed from the choice set.*
4. **Lack of Perfect Competition** *If R holds, and either Ua or Ub holds, but H does not hold, then in a symmetric Bertrand-Nash price setting equilibrium with single product firms, markups are almost surely bounded away from zero in the limit.*
5. **Contribution of Observed Characteristics** *If assumption R holds, and either Ua or Ub holds, and in addition the observable portion of utility is almost surely bounded above (there exists an $M < \infty$ such that for all (y_i, β_i), $|u(x_j, y_i - p_j, \beta_i)| < M$ a.s.) then the contribution of the observed characteristics to utility almost surely goes to zero as the number of products becomes large. (i.e., $\lim_{J\to\infty} \frac{\epsilon_{ij^*}}{u_{ij^*}} = 1$ a.s., where $j^* = \arg\max_{j \in 0..J} u_{ij}$).*
6. **Compensating Variation** *If R holds, and either Ua or Ub holds, then as the number of products becomes large the compensating variation for removing all of the inside goods almost surely tends to infinity for every individual.*

Property two, which holds necessarily in standard discrete choice econometric models, seems to contradict economic intuition. It seems unreasonable to allow $u(x_j, y_i - p_j, \beta_i)$ to head to negative infinity as products are added because that would imply either that the new products being added were particularly bad, or that individuals' taste coefficients are not constant when the choice set changes, neither of which makes sense.[9] Thus, we conclude that

[9] We note that this property has implications to the limiting behavior of the standard estimators for these models. If s_0 remains approximately constant for all market sizes, then the mean utilities for inside goods should change with the market size.

the probability of purchasing the outside good necessarily heads to zero for all individuals as the choice set becomes large.

If we are trying to describe choice behavior in a narrowly defined market, then it seems unreasonable that eventually all individuals would purchase some variety of the good just because there are many varieties available. There ought to be some characteristics that all inside goods share (e.g., all cell-phones have essentially the same use). If an individual has a strong negative taste for a common characteristic of the inside good (e.g., they have no cell-phone service in their area, or they simply do not like talking on the phone), then no matter how many varieties are available we should not necessarily expect the individual to purchase the good. At very least it would be desirable for the econometric model to be rich enough to allow for the possibility that the outside good retains positive share in the limit.

Properties three and four are closely related so we discuss them together. Property three implies that, even in the limiting case, the assumptions commonly maintained (e.g., in logit, GEV, and random coefficents logit models) are not sufficient to imply that individuals would be willing to switch to their second favorite product with zero compensation when the number of products becomes large. The technical reason for this is that product space may expand outwards too rapidly. As a result, markups may also tend to some positive number in the limit.[10] Again, if we are considering a narrowly defined market, then we might expect that the product space should fill up eventually and products should become close substitutes in the limit. Extreme value based models do not allow for this possibility.

Properties three and four do not hold necessarily, but depend on the shape of the distribution of ϵ_{ij}, and specifically the upper tails of the distribution. They hold for the GEV and the logit, including random coefficients logit models. This suggests that, even if independence is assumed, the probit model might have better economic properties than the logit model. In particular, probit may be preferrable to logit in certain applications such as welfare studies, where there may be a tendency for logit to overvalue additional choices, and studies of competition in differentiated products markets, where logit may tend to imply markups that are too high as a result of overestimating the differentiation between products. However, the practical importance of this result still needs to be investigated.

Property five shows that in standard discrete choice models the contribution to utility from observed variables changes depending on the number of products in the market, which also seems economically unintuitive. It seems more intuive that the percentage of the variance in utility explained by observables should remain more or less constant for any given market as the number of products in the market changes.

Property six singles out a problem with using discrete choice models for welfare analysis. The model implies that with enough products to choose from every individual will need arbitrarily large amounts of income to be as well off with the outside good alone as with the inside good. The implication is that every individual is costlessly receiving arbitrarily large (relative to

[10] Anderson, De Palma, and Thiesse show this for the logit model.

income or price) levels of utility from something about the good that we cannot observe.[11]

Note that all of the properties above are driven by the properties of the random error term. We now consider the dimension of the error term, where the term "dimension" loosely refers to the number of independent draws necessary to write down all of the elements of ϵ [Note: this argument is loose and is to be cleaned up in a future revision]. In the standard discrete choice econometric models, the dimension of the error term is $I * J$, where I is the number of individuals and J is the number of products. That is, the error term consists of $I * J$ independent draws from some distribution. Thus, if one product is added to the choice set, the dimension of the error term increases by I, which is typically a very large number (e.g., on the order of 100,000,000 in applications to demand for all U.S. households). Past empirical work has shown, not surprisingly, that under these conditions the undesirable properties above show up quite strongly in practice (see e.g. BLP, BLP2, Petrin (1998)).

The final property that we list is regarding the interpretation of such an error term. Past literature (Caplin and Nalebuff (1991)) has shown that the error term can be interpreted as a taste for products. The following shows that that is the only interpretation consistent with the assumptions made.

7. **Taste for Products** *The vector of error terms, ϵ, in the standard discrete choice econometric model has dimension equal to $I * J$, where I is the number of individuals and J is the number of products. Any error term consisting of unobserved product characteristics and unobserved individual tastes would have dimension at most $a * I + b * J$, where a is the number of unobserved tastes and b is the number of unobserved product characteristics. Because standard models imply that either a is a function of J or b is a function of I, the error term in standard models can only be interpreted as an individual level product-specific taste.*

In summary, the standard discrete choice econometric model have several properties which we think are undesirable from the point of view of economic theory. All of the properties 1-7 listed above are driven by the random error term in the utility function. Since this error term was included in the model primarily for econometric convenience rather than on theoretical economic grounds, in this paper we will proceed by considering choice models which do not rely on a random error term.

[11]Since the comparison is relative to income or price, this statement is regardless of normalization.

3 Model and Identification

3.1 A Pure Hedonic Model of Differentiated Products

In this section, we develop a model of consumer demand for differentiated products. Anderson, de Palma and Thisse (1992) identify three different types of models of demand for differentiated products — random utility models, representative agent models, and pure hedonic models.

In section 2, we demonstrated that random utility models have unappealing economic properties. Representative agent models are also not particularly appealing in many applications, both because the assumptions required for aggregation may not be plausible — even standard random coefficients discrete choice models do not generally have representative agent representations — and because they do not allow an investigation of the *distribution* of welfare effects from policy changes, our interest in this paper.

Thus, we believe that the most theoretically appealing model in many applications is the pure hedonic model. This section will describe a new approach to the identification and estimation of a pure hedonic discrete choice model.

3.2 The Pure Hedonic Model

In the pure hedonic model, a commodity is a collection of a finite number of attributes which we will represent as a vector of real numbers. For instance, if we were studying the automobiles market, then the attributes of an automobile relevant to a consumer might include horsepower, size, power steering and so forth. In this section, we will establish the basic notation and definitions for the pure hedonic model.

Let j represent a product and let J be the set of all products. The vector $(x_j, \xi_j) = (x_{j1}, ..., x_{jK}, \xi_j)$ will be interpreted as the attributes of product j. The first K attributes, which we denote by the vector x_j, are product attributes that are perfectly observed by both the consumer and the economist. The attribute ξ_j we will interpret as only observed by the consumer and not observed by the economist. We denote the set of product characteristics as $X \subseteq R^{K+1}$. Allowing for a product characteristic that is unobserved to the economist has been emphasized in the recent literature on demand estimation for differentiated products. Authors such as Berry (1994), BLP, BLP2, Nevo(2000), Petrin(1998), Davis(1999), among others, have emphasized that rarely does the economist observe all product characteristics, and that failing to take account of unobserved product characteristics may bias empirical demand estimates.

In our analysis, we assume that there are T markets. We let I_t be the set of all consumers in market $t \in T$ and we index a single consumer by $i \in I_t$. Let $I = \cup_t I_t$ be the set of consumers observed in all of the markets. If $I = I_t$ we observe all consumers in all markets whereas if $I_{t_1} \cap I_{t_2} = \{\phi\}$ for all $t_1 \neq t_2$ then we have only one observation per consumer.

Consumers are utility maximizers who select a product $j \in J$ along with a composite commodity $c \in R_+$. Each consumer i has a utility function $u_i(x_j, \xi_j, c) : X \times R_+ \to R$. The price of commodity j in market t is p_{jt} and the price of the composite commodity in market t is p_{ct}. Consumers have income y_i and consumer i's budget set in market t, $B(y_i, t)$, must satisfy:

$$B(y_i, t) = \{(j, c) \in J \times R_+ \text{ such that } p_{jt} + p_{ct}c \leq y_i\}$$

Consumer i in market t solves the following maximization problem:

$$\max_{(j,c) \in B(y_i,t)} \quad u_i(x_j, \xi_j, c) \tag{3}$$

We denote consumer i's demand correspondence as $\widetilde{h}(y_i, t)$, which is defined as:

$$\widetilde{h}(y_i, t) = \{(j, c) \in J \times R_+ : (j, c) \text{ solves } (3)\} \tag{4}$$

Definition. We say that $\widetilde{h}(y_i, t)$ is generated by $u_i(x, \xi, c)$ if $\widetilde{h}(y_i, t)$ satisfies (4).

Before we continue our exposition of the model, we show under weak conditions and using only demand based arguments that in any equilibrium, prices in each market have the following properties, (i) there is one price for each bundle of characteristics (that is, there is an equilibrium price surface), (ii) the price surface is increasing in the unobserved characteristic, and (iii) the price surface satisfies a Lipschitz condition.

Theorem 1. *Suppose that for every individual in every market,*

1. $u_i(x_j, \xi_j, c)$ *is continuously differentiable in* c *and strictly increasing in* c*, with* $\frac{\partial u_i(x_j,\xi_j,c)}{\partial c} > \epsilon$ *for some* $\epsilon > 0$ *and any* $c \in (0, y_i]$.
2. u_i *is Lipschitz continuous in* (x_j, ξ_j).
3. u_i *is strictly increasing in* ξ_j.

Then, for any two products j *and* j' *with positive demand in some market* t*,*

(i) *If* $x_j = x_{j'}$ *and* $\xi_j = \xi_{j'}$ *then* $p_{jt} = p_{j't}$.

(ii) *If* $x_j = x_{j'}$ *and* $\xi_j > \xi_{j'}$ *then* $p_{jt} > p_{j't}$.

(iii) $|p_{jt} - p_{j't}| \leq M(|x_j - x_{j'}| + |\xi_j - \xi_{j'}|)$ *for some* $M < \infty$.

Proof. See appendix. □

Before we continue the exposition of the model, some discussion of the theorem is in order. The intuition for the theorem is that if properties (i)-(iii) were not satisfied by the equilibrium prices, then some goods in the market could not have positive demand. Because the theorem is based on demand side arguments only, these results are general to many types of equilibria, both dynamic and static.

We use this theorem to justify the fact that we will remain largely agnostic on the supply side of the model in this paper. The lack of supply side information has benefits and costs. The benefit is that our results will be general to a large class of equilibria. The cost is that assumptions about supply would provide additional identifying information, specifically about the shape of the pricing function..

We also remind the reader that in section 2 we established that alternative choice models such as the standard discrete choice econometric models do not satisfy the conditions of the theorem (due to the error term) and hence, there does not necessarily exist a price surface for those models.

Because (iii) holds for all pairs of products, in the limit the price function will be Lipschitz continuous. Because differentiability is desirable in estimation, it would also be nice to establish a Lipschitz condition on the discrete analogs of the derivatives of the price function, so that in the limit the price function could be established as continuously differentiable. However, this is not possible to do using purely demand based arguments because a kinked price surface does not necessarily rule out positive demand everywhere. Differentiability of the price surface would necessarily come from both demand and supply side primitives, such as differentiability of preferences and of the cost function.[12]

Based on the results of the theorem, from this point onward we will take it for granted that there exists a continuous equilibrium price surface in each market relating the characteristics of the products in the market to their prices. We denote the equilibrium price surface for market t as $p_t(x_j, \xi_j)$. It is a map from the set of product characteristics to prices that satisifies $p_t(x_j, \xi_j) = p_{jt}$ for all $j \in J$ and we will assume that (i)-(iii) hold.

For clarification purposes, the price function in each market is an equilibrium function that is dependent upon market primitives. It does not tell us what the price of a good would be that is not already in the market. If a new good were added, in general all the prices of all the goods would change to a new equilibrium. It also does not tell us what the price of a good would be if any other market primitives were changed, such as consumer preferences, firm costs, or if the same good were to be produced by another multi-product firm. What it does tell us is the relationship between characteristics and prices as perceived by a consumer in this particular market.

[12]For example, if the cost function were continuously differentiable and the market were perfectly competitive then the price function would be continuously differentiable.

3.3 Identification: Discussion

While empirical economists have recently stressed the importance of accounting for unobserved product attributes when estimating demand for differentiated products, there has been, to the best of our knowledge, no formal development of integrability for differentiated products when the economist cannot observe all of the product characteristics. When all of the product characteristics are observed, this is a standard problem and the assumptions under which the underlying preferences can be recovered are well understood. Varian (1988) shows that in the case where the price function is linear and one good is not observed, revealed preference arguments provide essentially no information about preferences. In the next section (3.4) we similarly establish that, in general, it is not possible to recover consumer i's weak preference relation even if the consumer's entire demand function is known. The intuition behind the result is straightfoward. Without any restrictions, it is possible to attribute all of the variation in price to the unobserved product characteristic ξ. Since the observed choices satisfy the strong axiom of revealed preference, it is possible to construct a utility function u_i, that only depends on ξ and the composite commodity, that will generate the observed demands.

In sections 3.5 and 3.6, we discuss two sets of conditions under which the equilibrium price function in all markets t, $p_t(x_j, \xi_j)$, and the unobserved product characteristics $\{\xi_j\}$ are identified. The first set of conditions we discuss rely on independence. Specifically, we assume that the unobserved product characteristic ξ_j is independent of the observed product characteristics x_j. We also allow that prices are measured with (an additively separable) error. We generalize the arguments of Matzkin (1999) to demonstrate that the price functions, $p_t(x_j, \xi_j)$, and the unobservable characteristics, $\{\xi_j\}$, are identified.

The second set of conditions we refer to as the "options package approach". In many product markets, such as automobiles or computers, the consumer's choice can be broken down into first choosing a model (such as a Camry, Corolla, Taurus...), and second choosing an options package (horsepower, air conditioning, power steering and so forth). We establish that if the ξ_j is isomorphic to the "model" and certain regularity conditions hold, then the price functions $p_t(x_j, \xi_j)$, and the unobservable characteristics, $\{\xi_j\}$, are identified.

Once the price surface and unobservables are known, it is possible to uniquely recover (under regularity conditions) the preferences that generated the consumer's demand correspondence. The results of the next section establish, in an idealized setting and with a sufficiently rich set of observations on individual behavior, that it is possible to recover the consumer's preferences for both observed and unobserved product characteristics.

3.4 Non-Identification Result

We begin with the following assumptions about the primitives of our model.

A1. $J = [0, 1]$

A2. X is a convex, compact subset of R^{K+1} and $0 \in X$.

A3. For all t, $p_t(x, \xi)$ is a continuous function, $p_t(0) = 0$ and $p_{ct} > 0$.

A4. $u_i(x, \xi)$ is continuous.

The first assumption A1 implies that there are a continuum of products in our model. We have chosen, for the sake of conserving on notation, to assume that the set of products is the unit interval. Assumptions A2 and A3 are made in order to insure that the budget set is bounded and assumption A4 will guarantee that the demand correspondence is non-empty. We formalize these observations in the Theorem below.

Theorem 2. *Assume A1-A4 hold. For all $y_i > 0$ and for all t, $B(y_i, t)$ is a compact set. For all t and all $y_i > 0$, $\tilde{h}(y_i, t)$ is non-empty.*

Proof. Since the pricing function $p_t(x, \xi)$ is continuous and $p(0) = 0$, the budget set is closed and non-empty. Since X is bounded, the budget set is compact. Since utility is a continuous function, there is at least one utility maximizing bundle. Therefore the demand correspondence is not empty. □

In revealed preference, it is often convenient to work with preference relations instead of utility functions since utility functions are never uniquely determined. Below, we define a weak preference relation for consumer i.

Definition 1. *We will say that $\succeq_i$ is a **weak preference relation** for consumer i if for all $j, j' \in J$, $x_j \succeq_i x_{j'}$ if and only if $u_i(x_j, \xi_j) \geq u_i(x_{j'}, \xi_{j'})$.*

Note that given a utility function u_i there is a unique binary relation $\succeq_i$ that is a weak preference relation for our consumer.

Definition 2. *We say that j is **directly revealed preferred** by i to j' if there exists an income level y_i and a market t such that $j, j' \in B(y_i, t)$, $j \in \widetilde{h}(y_i, t)$ and $j' \notin \widetilde{h}(y_i, t)$. If j is revealed prefered to j' we write jS_ij'.*

Definition 3. *We say that S_i satisfies the **strong axiom of revealed preference** if S_i is acyclic, that is, there does not exist $j_1, j_2, ..., j_n$ such that:*

$$j_1 S_i j_2 \text{ and } j_2 S_i j_3 \text{ and...and } j_{n-1} S_i j_n \text{ and } j_n S_i j_1$$

In our pure hedonic model, as in the standard analysis of the consumer in partial equilibrium, maximization will imply that the strong axiom of revealed preference is satisfied.

Theorem 3. *If $\widetilde{h}(y_i, t)$ is generated by $u_i(x, \xi, c)$ then $\widetilde{h}(y_i, t)$ satisfies the strong axiom of revealed preference.*

Proof. Standard. □

We now turn to the problem of identification. That is, we wish to know whether those objects that are typically not observable to the economist, such as the hedonic pricing function, unobserved product characterisitc and weak preference relation are uniquely determined by those objects the economist might typically expect to observe in an empirical study, the first K product characteristics, prices and consumer choice. We define identification formally below.

Definition 4. *We say that the pricing functions $\{p_t\}_{t\in T}$ and unobserved product characteristics $\{\xi_j\}$ are **identified** if*

a. For all $t \in T$ and for all j, $p_t(x_j, \xi_j) = p_{jt}$.

b. If $\{\widetilde{p}_t\}_{t\in T}$ and $\{\widetilde{\xi}_j\}$ satisfy (a), then $\{\widetilde{p}_t\}_{t\in T} = \{p_t\}_{t\in T}$ and $\{\widetilde{\xi}_j\} = \{\xi_j\}$.

Definition 5. *We say that the weak preference relation $\succeq_i$ is identified if*

a. $\succeq_i$ generates the demand correspondence $\widetilde{h}(y_i, t)$ and

b. If any other weak preference $\widetilde{\succeq}_i$ generates the demand correspondence $\widetilde{h}(y_i, t)$ then $\widetilde{\succeq}_i = \succeq_i$.

We begin by showing that without further assumptions, neither the hedonic pricing function nor the weak preference relation is identified. This result will hold so long as there is at least one point at which the price function and the utility function are both strictly increasing in the observed product characteristics.

A5 There exists at least one $\overline{x} \in X$ and at least one market t such that $p_t(x)$ is strictly increasing in its first k arguments in some neighborhood of $\overline{x}$.

A6 There exists at least one point $(\overline{x}, \overline{c})$ such that $u_i(x, c)$ is strictly increasing in some neighborhood of $(\overline{x}, \overline{c})$.

Theorem 4. *Suppose that A1-A5 hold. Then $\{p_t\}_{t\in T}$ and $\{\xi_j\}_{j\in J}$ are not identified.*

Proof. Set $\xi_j = j$. Define the price function to satisfy $p_t(x_j, \xi_j) = p_{jt}$ for all x_j. Since p_t is nowhere strictly increasing in the observed product characteristics the hedonic pricing function is not identified. □

Theorem 5. *Suppose that the demand correspondence $\widetilde{h}(y_i, t)$ is generated by the utility function u_i. Also suppose that A1-A4 and A6 hold. Then the weak preference relation $\succeq_i$ is not identified.*

Proof. Using the construction in the previous theorem, suppose that all of the price is due to the product unobservable. Since the demand correspondence is generated by a utility function it will satisfy the strong axiom of revealed preference. Since demand obeys the strong axiom of revealed preference, it is possible, following standard arguments provided in Richter(1966), to construct a preference relation only over the unobserved product characteristic ξ_j that generates the observed demand. It is then trivial to show this preference relation is nowhere strictly increasing in the observed characteristics. This violates assumption A6. A complete proof is provided in the appendix. □

The above theorems demonstrate that if the economist fails to perfectly observe all product characteristics then it is not possible to identify the hedonic pricing function or the consumer's weak preference relation. This result stands in sharp contrast to standard results in revealed preference which show that observed choice behavior is enough to recover a consumer's weak preference relation. Outside of experimental settings, it is seldom possible for the economist to observe all of the product characteristics. We believe, therefore, that it is important to investigate whether there are *any* conditions under which it is possible to recover the consumer's weak preference relation using information that might plausibly be available to the economist in an empirical study.

3.5 Identification Using Independence.

In this section we demonstrate that the price function and the unobservables $\{\xi_j\}$ are identified if the unobserved product characteristic ξ is independent of the observed product characteristics x. This is true even if the econometrician observes price with measurement error.

We first consider identification of the price surface in the case where there is no measurement error. If there is no measurement error, then the observed prices are equal to the equilibrium price surface,

$$p_{jt} = p_t(x_j, \xi_j), \tag{5}$$

where $p_t : A \times E \to \mathbb{R}$, $A \subseteq \mathbb{R}^K$ is the support of x, and $E \subseteq \mathbb{R}$ is the support of ξ. For the case where there is a single market and no measurement error in prices, Matzkin (1999) shows under weak conditions that both the functional form of $p_1(\cdot)$, and the distribution of the unobserved product characteristics, $\{\xi_j\}$, are identified up to a normalization on ξ. Thus, the first part of our identification proof follows that of Matzkin (1999), the only difference being that we extend her results to cover the case of many markets. We begin with two assumptions,

B1 ξ is independent of x.

B2 For all markets t and all x, $p_t(x,\cdot)$ is strictly increasing, with $\frac{\partial p_t(x,\xi)}{\partial \xi} > \delta$ for all (x,ξ) for all t and some $\delta > 0$.

Assumptions B1 and B2 are the primary identifying assumptions. B2 ensures that at each x there could only be one value of the unobservable consistent with each price. The lower bound on the derivative is needed to ensure that as the number of markets becomes large the price function does not become arbitrarily close to a weakly increasing function.

Let the set I be the set of price functions satisfying B2, and Γ be the set of distribution functions that are strictly increasing.

$$I = \{p' : A \times E \to \mathbb{R} \mid \text{for all } x \in X,\ p'(x,\cdot) \text{ is strictly increasing}\} \tag{6}$$
$$\Gamma = \{F : \mathbb{R} \to \mathbb{R} \mid F \text{ is strictly increasing}\} \tag{7}$$

Since the unobserved product attribute has no inherent units, it will only be possible to identify it up to a normalization. Thus, without loss of generality, we assume that a normalization has been made to ξ such that at some point $\bar{x} \in X$ the pricing function in one market is equal to the unobservable, ξ. Because the price function is monotonic in ξ, this normalization amounts to a monotonic transformation of ξ and the price function. We define the set of functions characterized by this normalization as,

$$M = \{p' : A \times E \to \mathbb{R} \mid p' \in I \text{ and } p'(\bar{x},\xi) = \xi\} \tag{8}$$

In the theorem below, we will assume without loss of generality that $p_1(\cdot) \in M$.[13] Define identification to mean identification within the set satisfying the normalization made above,

Definition 6. *The pair (p_1, F_ξ) is **identified in** $(M \times \Gamma)$ if*

i. $(p_1, F_\xi) \in (M \times \Gamma)$, *and*

ii. For all $(p', F'_\xi) \in (M \times \Gamma)$,

$$[F_{p,x}(\cdot; p, F_\xi) = F_{p,x}(\cdot; p', F'_\xi)] \Rightarrow [(p, F_\xi) = (p', F'_\xi)]$$

We now show that identification holds in the case where there is no measurement error.

Theorem 6. *If prices are observed without error, B1-B2 hold, and if $p_1 \in M$, then (p_1, F_ξ) is identified in $(M \times \Gamma)$, and p_t is identified in I for all $t > 1$. Furthermore, $\{\xi_j\}$ is identified.*

[13] The set M here corresponds to the set M2 in Matzkin (1999).

Proof. Identification of p_1, F_ξ, holds by Matzkin (1999) Theorem 1. Identification of the price function in the remaining markets is as follows,

$$p_t(x_0, e_0) = F^{-1}_{p_t|x=x_0}(F_\xi(e_0)), \tag{9}$$

where $F_{p_t|x=x_0}$ is the observed distribution of prices in market t at the point x_0.

To show that the unobserved product characteristics are identified, note that for each product j,

$$\xi_j = F_\xi^{-1}(F_{p_1|x=x_j}(p_{j1})) \tag{10}$$

□

From the proof of the theorem we can see that if there is no measurement error, cross-market variation is not needed for identification of the unobserved product characteristic.

We now consider the case where prices are observed with measurement error that is independent of both x and ξ. Specifically, we assume that p_{jt} is not observed. Instead, the econometrician observes y_{jt}, where

$$y_{jt} = p_{jt} + \epsilon_{jt} \equiv p_t(x_j, \xi_j) + \epsilon_{jt} \tag{11}$$

For this case, some additional assumptions regarding the measurement error are necessary.

B3 ϵ_{jt} is *iid*, and $E[\epsilon|x, \xi] = 0$.

These are standard assumptions regarding measurement error. However, we note that for the purposes of identification it is not necessary that ϵ_{jt} be *iid*. All that matters is that, for every x and ξ, a law of large numbers holds for ϵ_{jt} across each of j and t.

We now show identification for the measurement error case.

Theorem 7. *If prices are observed with error, B1-B3 hold, and if* $p_1 \in M$, *then* (p_1, F_ξ, F_ϵ) *is identified in* $(M \times \Gamma \times \Gamma)$, *and* p_t *is identified in* I *for all* $t > 1$. *Furthermore,* $\{\xi_j\}$ *is identified.*

Proof. Let

$$\bar{p}^T(x, \xi) = \frac{1}{T}\sum_{t=1}^{T} p_t(x, \xi) \tag{12}$$

and let $\bar{p}_j^T \equiv \bar{p}^T(x_j, \xi_j)$. For each product, j, we can observe $\bar{p}_j^T$ by averaging the observed prices, y_{jt}, across markets. Since the measurement error is conditional mean zero for every (x, ξ), it averages to zero for large T.[14]

[14]Note that we do not assume that $\lim_{T\to\infty} \bar{p}^T(x_j, \xi_j)$ exists.

For each product, j, define the set

$$J_j = \{k \in J|\ x_k = x_j \text{ and } \lim_{T\to\infty} \bar{p}_j^T - \bar{p}_k^T = 0\} \tag{13}$$

The set J_j is the set of all products with the same characteristics, both observed and unobserved, as product j. The value of the price function for each product j, p_{jt} is identified by averaging prices within each market t across the set of products J_j:

$$p_{jt} = E[y_{kt}|k \in J_j] \tag{14}$$

The measurement error again averages to zero.

Since the value of the price function is identified for each product in each market, the rest of the proof of identification for F_ξ, $\{\xi_j\}$, and $p_t(\cdot)$ follows by Theorem 6 above.

Finally, $\epsilon_{jt} = y_{jt} - p_t(x_j, \xi_j)$, so ϵ_{jt} and the joint distribution of ϵ and x and ξ is also identified. □

These results demonstrate that the price functions $p_t(x, \xi)$ and the unobserved product characteristics $\{\xi_j\}$ are identified (up to a normalization). Once ξ is known, recovering the weak preference relation $\succeq_i$ from observed choice behavior is the standard integrability problem. It is well understood that there exist fairly weak regularity conditions under which $\succeq_i$ can be uniquely recovered from the data. See Hurwics and Uzawa (1971) for a discussion.[15]

3.6 Identification Using "Options Packages"

We believe the independence assumption made in the previous subsection to be quite strong. Therefore, in this section we provide a weaker set of assumptions which also provide identification and which we believe may be satisfied in many applications of interest.

In many applications, consumers may be simultanesouly choosing a model and an options package for that model. For instance, a car buyer's problem could be represented as choosing a model (Camry, Taurus, RAV4,...) and a package of options associated with the model (horsepower, air conditioning, power steering, ...). Purchases of computers might also be well represented as the joint choice of a model (Dell Dimension 8100, Gateway Profile 2, Compaq Presario 5000 Series,...) and an options package (amount of RAM, type of processor, hard drive size,...). In this section, we demonstrate that if it is the case that the product

[15] Since it is sometimes the case that the econometrician does not observe all products in all markets, the question of identification of the unobservable for products not observed in the reference market (market 1) may be of interest. Identification of the unobservable for products which are not observed in the reference market can be obtained by comparing their average prices across markets with the average prices of products which are observed in the reference market. As long as the products of interest are observed in countably many markets simultaneously with other products that are observed in the reference market, identification is still obtained (using the cross-market average prices as in the proof above).

unobservable ξ_j corresponds to a model and the x_j correspond to an options package then it is possible to identify the pricing function and the weak preference relation.

We begin by providing a precise definition of what it means to be a model. For the purposes of our analysis, the set of models will be a partition of J. We let z denote a model and Z denote the set of all models. We assume that there exists a map $\pi : J \to Z$ which associates products with models. The inverse image of z under the map π are those products which are model z, although with possibly different options packages. We assume that z is observable, and that x and z have joint distribution $F_{x,z} : A \times Z \to \mathbb{R}$.

The first assumption in this section is that all products which are the same model, have the same value of the unobservable,

B4. For all $j_1, j_2 \in Z$, if $\pi(j_1) = \pi(j_2)$ then $\xi_{j_1} = \xi_{j_2}$.

In order to identify the product unobservable, we need there to be a "baseline" or standard options package that is available for all models z. We formalize this requirement using the following assumption,

B5. There exists an $\bar{x} \in A$ such that for all $z \in Z$, $f(\bar{x}|z) > 0$.

The baseline package here corresponds to the "reference" package in the above section. Due to the lack of implicit units for ξ, we will again only be able to identify ξ and the price function up to a normalization. In this case we assume that the price function in market 1 has been normalized such that at the baseline package, $\bar{x}$, it equals the unobservable, ξ.

Finally, let

$$\Gamma' = \{F : \mathbb{R}^2 \to \mathbb{R} \mid F \text{ is strictly increasing in the natural ordering of } \mathbb{R}^2\} \tag{15}$$

We are now ready to show identification, again beginning with the case where there is no measurement error,

Theorem 8. *If prices are observed without error, B2 and B4-5 hold, and if $p_1 \in M$, then $(p_1, F_{x,\xi})$ is identified in $M \times \Gamma'$, and p_t is identified in I for all $t > 1$. Furthermore, $\{\xi_j\}$ is identified.*

Proof. For each product, j,

$$\begin{aligned} \xi_j &= p_1(\bar{x}, \xi_j) && (16) \\ &= p_{1k} \text{ for } k \in \pi^{-1}(\pi(j)) \text{ such that } x_k = \bar{x}. && (17) \end{aligned}$$

Equation (16) holds due to the normalization. A baseline product, k, exists for every model $\pi(j)$ by B5. This equation identifies $\{\xi_j\}$ and $F_{x,\xi}$.

The price function in each market is then given by the prices of non-baseline packages. For any point $(x_0, e_0) \in A \times E$,

$$p_t(x_0, e_0) = p_{kt} \text{ for } k \in J \text{ such that } \xi_k = e_0 \text{ and } x_k = x_0 \tag{18}$$

□

Proving identification when there is measurement error in prices is trivial since models are observed.

Theorem 9. *If prices are observed with error, B2-3 and B4-5 hold, and if $p_1 \in M$, then $(p_1, F_{x,\xi}, F_\epsilon)$ is identified in $(M \times \Gamma' \times \Gamma)$, and p_t is identified in I for all $t > 1$. Furthermore, $\{\xi_j\}$ is identified.*

Proof. Let $J_j = \{k \in \pi^{-1}(\pi(j)) \mid x_k = x_j\}$. As above, J_j is the set of all products with the same characteristics as j. Then

$$p_t(x_j, \xi_j) = E[y_{kt} | k \in J_j], \tag{19}$$

where the measurement error again averages to zero. The rest of the proof is by Theorem 8. □

We note that, unlike the independence case above, in this case cross-market variation in prices is not needed for identification. We instead use the fact that models are observed to group products according to their unobservable. However, as shown in the independence case, cross-market variation in prices would provide us with an additional source of identification for the unobservable. This is important because in estimating the model it would be optimal to use both sources of information.

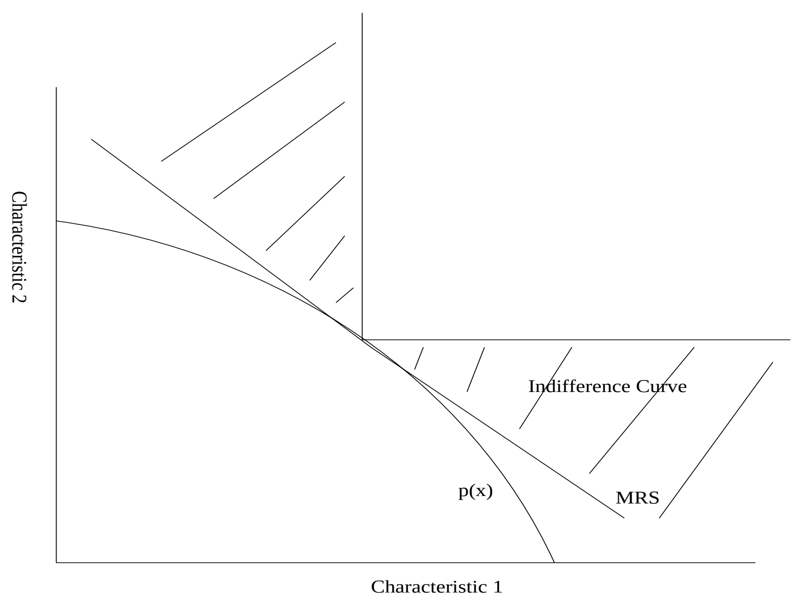
These results demonstrate that the price functions $p_t(x, \xi)$ and $\{\xi_j\}$ are identified up to a normalization from the observed demand behavior. As above, once ξ is known, recovering the weak preference relation $\succeq_i$ from observed choice behavior is the standard integrability problem.

3.7 Identification of Preferences with a Continuous Choice Set and a Finite Number of Observations Per Individual.

In the previous section, we studied the integrability problem in the case where the economist can observe the entire demand function. It is important to study this case to know, in principle, as the number of observations per individual becomes sufficiently large that we will be able to identify the consumer's preferences. However, in applied work the econometrician is typically faced with a situation where he observes only a handful of choices per consumer, often just one. In this section, we study identification when the economist only observes the individual choose a finite number of times.

When the entire demand function is not observed, it is clear that recovery of the entire weak preference relation is not possible. In figure 3.7 we suppose that the good is a bundle of two characteristics. The slope of $p(x)$ at the budget line identifies the marginal utilities at the chosen bundle. Without further assumptions, we can learn nothing more. However, some weak assumptions (we use these only for illustrative purposes) can tell us a range within which the indifference curve must lie. If we assume a diminishing MRS, then the indifference curve must lie everywhere above the tangency line at the chosen bundle, providing us a global lower bound on the indifference curve. If we assume monotonicity of preferences, then the indifference curve must lie everywhere below the indifference curve for Leontieff preferences. Together, the two assumptions allow us to conclude that the indifference curve must lie in the shaded area of our figure.

Figure 1: Global Identification of Indifference Curves



Even with these two assumptions, we are still left with a wide range of possibilities. One way to narrow down the range of possibilities is to use functional form assumptions for the utility function. Many discrete choice models used in the previous literature specify the utility function as being linear in product characteristics. In that case, the random coefficients *are* the marginal utilities. Thus, for this commonly used case, looking at the tangency conditions for all consumers in the population allows non-parametric identification of the population distribution of random coefficients (conditional on the functional form of the utility function) even if individuals are only observed once. Identification of the indifference curve away from the point of tangency is based on the functional form of utility.

We now derive identification conditions for a general parametric model of preferences. An agent in this economy is characterized by a B dimensional parameter vector β_i that is an element of R^B. We write the utility function as:

$$u_i(x, c) = u(x, y_i - p(x); \beta_i). \tag{20}$$

Agents are assumed to choose the element $x \in X$ that maximizes utility. We note that since the previous section has shown that the unobservables, $\{\xi_j\}$, are identified by the price function, we proceed in this section as if all characteristics are observed.

If both u and $p(x)$ are differentiable in x and if agents choose an interior maximizer, then the following first order conditions are necessary for maximization:

$$\frac{\partial}{\partial x_k} \{u(x, y_i - p(x); \beta_i)\} = 0 \text{ for } k = 1, ..., K \tag{21}$$

The intuition behind the identification and estimation of our model is similar to the two-stage approach of the earlier literature in hedonics. If consumers are maximizing utility, then it must be the case that the slope of the indifference line is tangent to the budget set. Therefore, if the econometrician can consistently estimate $p(x)$ as a function of characteristics that are both observed and unobserved to the econometrician, then it may be possible to recover a consumer's marginal utilities from standard tangency conditions.

The next theorem derives formal conditions under which individual preferences are identified by structural assumptions when individuals are only observed making a finite number of choices. For simplicity, we consider only the case where individuals are only observed a single time. The case where the consumer is observed more than once is a straightforward extension.

We assume that the econometrician observes the choices $j = 1..J$ of a randomly sampled group of $i = 1..I$ individuals, in a single market. Suppose that agent i is consuming a product defined by a vector of characteristics x. The Jacobian for the first order conditions (21) for agent i is:

$$D_\beta \left(D_{x'} \{u(x, y_i - p(x); \beta_i)\}\right) \tag{22}$$

Theorem 10. *Suppose $\beta_i \in \mathbb{R}^B$ and $x \in \mathbb{R}^K$. Then if the rank of the $K \times B$ matrix given by (22) is greater than or equal to B for all bundles x, then β_i can be written as a function of the consumption bundle x.*

Proof. This follows directly from the Global Inverse Function Theorem. □

Based on the above theorem, checking whether or not a particular utility function is identified is a straightforward application of counting equations and unknowns and applying the basic techniques of analysis. For any random coefficient model under consideration, a sufficient condition to prove identification is to verify whether the Jacobian(22) is invertible. This can often be handled in a straight forward fashion on a case by case basis.[16]

Theorem 10 places tight restrictions on the types of utility functions that can be identified using the choice data. Conditional on knowing the price surface $p(x)$, a single observation on an individual tells us only that individual's marginal utilities for each characteristic x_k at a single point in characteristic space. It should be obvious that this allows us to identify at most K random coefficients, and even in the case where there are only K random coefficients the Theorem places tight restrictions on the functional forms for the utility function that can be estimated.

One interpretation of this result is that only a first order approximation of the utility function can be identified (for each individual) using data containing only a single choice per individual. However, a first order approximation to the utility function should provide accurate results if the researcher is considering only local changes in utility for each individual. For example, the experiment of removing a single good from the market (to evaluate welfare obtained from the good) would involve only local changes to utility if the choice set is rich enough because individuals who previously purchased the good removed would in general substitute to goods with very similar characteristics.

However, we emphasize that it is not necessary to estimate a utility function that is linear in the characteristics. For counterfactuals relying on the global properties of indifference curves other more reasonable functional forms such as Cobb-Douglas (log-linear) may be preferred. Such assumptions should be carefully considered because they are untestable and impact the results. However, there is a wealth of economic theory that can be tapped for choosing the functional form for $u(\cdot)$.

In such cases, it would make sense to perform robustness checks on the results of interest subject to changes in the functional form of $u(\cdot)$. As above, at the individual level linear utility and Leontieff utility represent two extremes on the substitutability of characteristics. Robustness to these two alternatives may be prudent. The previous literature on discrete choice has focused on the case of linear utility, which allows for an extreme amount of substitutability. Log-linear utility, for example, may provide a better intermediate case.

While the theorem does place restrictions on what functional forms can be identified, we emphasize that it places no restrictions on the relationship between taste coefficients. In general, the correlations between taste coefficients are identified.

[16]If the consumer is only observed once, then it will often be possible to verify identification by simple inspection. If the consumer is observed more than once then verification of the identification conditions is slightly more difficult.

3.8 Identification of Preferences if the Choice Set is Discrete.

In practice, there are at least three reasons why the continuous choice model might not provide a good approximation to choice behavior. First (1), the number of products in the choice set may not be sufficiently large that the choice set is approximately continuous. If a consumer has only a handful of choices available to her then her observed choice may be very far from the bundle of characteristics that would maximize her utility simply because the latter is not available. Second (2), many product characteristics are fundamentally discrete. While miles per gallon and fuel efficiency might naturally take on continuous values, power steering and airbags are better represented by binary variables. Third (3), it may not be possible to reliably infer marginal utilities for consumers who choose products at boundaries. For example, while the commodity space for personal computers is quite dense in the interior of the space, for consumers that buy products on the boundary of this space (e.g., the fastest CPU currently available) there is a corner solution to their utility maximization problem and we can not reliably infer their taste coefficients.

If the product space is discrete, then in place of the marginal conditions in (21) we can only derive a set of inequality constraints. That is, if consumer i chooses product $j \in 1..J$ then it must be the case that

$$u_{ij}(\beta_i, x_j, y_i - p_j) \geq u_{ik}(\beta_i, x_k, y_i - p_k) \text{ for all } k \neq j \tag{23}$$

Therefore, it must be that $\beta_i \in A_j$, where

$$A_{ij} = \{\beta_i : u_{ij}(\beta_i, x_j, y_i - p_j) \geq u_{ik}(\beta_i, x_k, y_i - p_k) \ \forall k \neq j\}. \tag{24}$$

Thus, we have the result that if the choice set is discrete then the β_i parameters are not identified, meaning that we can not learn their exact values. However, that does not mean that the data is non-informative as to the taste coefficients. If the choice set is rich, the A_j sets may be very small. We show in appendix section 6.4 that if all of the characteristics are continuous and the choice set is compact, then as the number of products increases, the A_{ij} sets converge to the individual taste coefficients β_i (where it is assumed that individual i purchases good j).

In applications where the A_{ij} sets are large enough that the lack of identification matters, it is possible to proceed in two ways. First, we could use the A_{ij} sets to construct bounds on the aggregate distribution of the taste coefficients. Second, it is possible to use Bayesian techniques to identify one candidate aggregate distribution of interest. In section 4.3, we follow the latter strategy. In either case, the identification of the aggregate distribution is weakest in the cases of (2) and (3) above. By the results of section 6.4, as the number of products increases we get perfect identification of the distribution in the case of (1) above.

As in the identification arguments made above and earlier, in the sections that follow we propose a two-stage estimation strategy that is similar in spirit to the two-stage hedonics approach of Rosen (1974), Epple (1987), and Bartik (1987).

4 Estimation

4.1 Estimation, Stage 1: Unobservable Characteristics and the Price Surface – Independence Case

We assume that the econometrician observes prices and characteristics for $j = 1..J$ products across $t = 1..T$ markets. In this section we will maintain all of the assumptions in section 3.5. In particular, we assume that x, ξ, and ϵ are jointly independent. We leave out estimation of the options packages case here for the sake of brevity.

If there is no measurement error, then the estimators needed for our first stage are provided by Matzkin (1999). In the discrete choice set case (section 4.3 below) our first stage consists of using prices to estimate the value of the unobservables. We show how to use the Matzkin (1999) estimators to estimate the unobservables in section 4.1.2 below. In the continuous choice set case, it is also necessary to know the price function derivatives, for which Matzkin (1999) also provides estimators. If there is measurement error, then before using the Matzkin estimators it is necessary to do some smoothing to remove the measurement error. We show how to do this using a kernel estimator in the following subsection.

4.1.1 Removing the Measurement Error

If there is measurement error in prices, then the first step of the estimation procedure is to estimate the following general relationship,

$$y_{jt} = p_t(x_j, \xi_j) + \epsilon_{jt} \tag{25}$$

using the expectation in equation (14). To do this, as in the identification section, we will utilize average prices of products across markets.

Note that for each T, $\bar{p}^T(x, \cdot) \equiv \frac{1}{T}\sum_{t=1}^{T} p_t(x, \cdot)$ is strictly increasing for every x since each $p_t(x, \cdot)$ is. Thus, we can invert it,

$$\xi = \Psi(x, \bar{p}^T) \tag{26}$$

Let

$$g_t^T(x, \bar{p}^T) \equiv p_t(x, \Psi(x, \bar{p}^T)) \tag{27}$$

Then, by definition,

$$g_t^T(x_j, \bar{p}^T(x_j, \xi_j)) \equiv p_t(x_j, \Phi(x_j, \bar{p}^T(x_j, \xi_j)) \equiv p_t(x_j, \xi_j) \equiv p_{jt} \tag{28}$$

We will estimate the true prices, p_{jt} using a nonparametric estimator of $g_t^T()$. Within each market, t, a smooth kernel estimator for the true prices p_{jt} is given by $\hat{g}_t^T(x_j, \bar{y}_j^T)$, where

$$\hat{g}_t^T(x_0, \bar{p}_0^T) = \frac{\frac{1}{J}\sum_{k=1}^{J} \frac{1}{h_J^{K+1}} K_1\left(\frac{\bar{p}_0^T - \bar{y}_k^T}{h_J}\right) K_2\left(\frac{x_0 - x_k}{h_J}\right) y_{kt}}{\frac{1}{J}\sum_{k=1}^{J} \frac{1}{h_J^{K+1}} K_1\left(\frac{\bar{p}_0^T - \bar{y}_k^T}{h_J}\right) K_2\left(\frac{x_0 - x_k}{h_J}\right)}. \tag{29}$$

$\bar{y}_k^T = \frac{1}{T}\sum_{t=1}^{T} y_{kt}$ is an estimator for $\bar{p}_k^T$, which we need to estimate because it is not observed directly due to the measurement error. This will affect the finite sample performance of the estimator, but not the asymptotic performance so long as T grows fast enough with J.

We make the following assumptions (details in appendix):

C1 $K_1(\cdot)$ and $K_2(\cdot)$ are bounded real-valued Borel measurable functions satisfying $\int K(r)dr = 1$, and $K_1(\cdot) \in \mathcal{K}_{1,m}$ and $K_2(\cdot) \in \mathcal{K}_{K,m_1}$ for some $m_1 \geq 2$. $K_1(\cdot)$ has continuous derivatives up to order m_2.

C2 $\lim_{J\to\infty} h_J = 0$, $\lim_{J\to\infty} J h_J^{K+1} = \infty$, $\lim_{J\to\infty} h_J^m \sqrt{J h_J^{K+1}} = \lambda$ where $0 \leq \lambda < \infty$.

C3 x_j, ξ_j are *iid*, mutually independent, and distributed according to $F(x, \xi)$, with density $f(x_j, \xi_j)$.

C4 The functions $f(x_j, \xi_j)$ and $g_t^T(x_j, \bar{p}^T(x_j, \xi_j)) f(x_j, \xi_j)$ belong to $\mathcal{D}_{K+1,m_1}$ for all t, T.

C5 ϵ_{jt} is *iid*, and independent of x and ξ, $E[\epsilon_{jt}] = 0$, $E[\epsilon_{jt}^2|x_j] = \sigma^2(x_j)$, and $E|\epsilon_{jt}|^r < \infty$ for some $2 < r \leq \infty$.

C6 $J^{-2/r} T h_J^{\frac{2(m_2+1)}{m_2}} \to \infty$, where r is as in C5.

The non-standard assumptions are C5 and C6, which are needed in order to assure that the estimated $\bar{y}_k^T$ terms do not affect the estimation of the true prices. C6 requires that T increases fast enough with J, but T can still increase much slower than J, with the exact speed depending on the dimension of the problem, K, the properties of the measurement error distribution, and the smoothness of $K_1(\cdot)$. If the measurement error is either bounded or Normally distributed ($r = \infty$), and $K_1(\cdot)$ is very smooth, then T can increase slowly with J.

Theorem 11. *Under C1-C6,*

(i) $\sup_{\{(x_j,\xi_j)\in \mathbb{R}^{K+1}: f(x_j,\xi_j)>\delta\}} |\hat{g}_t^T(x_j, \bar{y}_j^T) - p_t(x_j, \xi_j)| \to 0$ *in probability.*

(ii) For all (x_j, ξ_j), $\sqrt{J h_J^{K+1}}(\hat{g}_t^T(x_j, \bar{y}_j^T) - p_t(x_j, \xi_j)) \to N\left(\frac{\lambda b(x_j, \bar{p}^T(x_j,\xi_j))}{f(x_j,\xi_j)},\ \frac{\sigma_\epsilon^2(x_j,\xi_j)}{f(x_j,\xi_j)} \int K(r)^2 dr\right)$

4.1.2 Estimation of $\{\xi_j\}$

Let $\hat{F}_{p_1|x=x_0}(e_0)$ be an estimator for the conditional distribution of prices given $x = x_0$ at the point e_0. For example, if there is no measurement error in prices, then the estimators outlined in Matzkin (1999) could be used. Define an estimator for ξ by the following,

$$\hat{\xi}_j = \hat{F}^{-1}_{p_1|x=\bar{x}} \hat{F}_{p_1|x=x_j}(\hat{p}_{j1})) \tag{30}$$

While Matzkin (1999) does not explicitly consider estimation of the unobservable, the asymptotic properties of the estimator in (30) are analogous to those of the estimator considered in Theorem 4 of that paper.

If there is measurement error, then the same estimator can be used except that it is first necessary to estimate the true prices as in section 4.1.1 above. Note that after plugging in the estimated true prices, the asymptotic properties of the estimator will change. This is because the estimator in section 4.1.1 has dimension $K + 1$ while the estimator $\hat{F}$ has dimension K. Again for brevity and because much of the work would replicate the results in Matzkin (1999), we omit the asymptotic properties of the measurement error estimator here.

4.2 Estimation of Preferences, Continuous Case

In this section we outline a strategy for estimating preferences for the case of one observation per individual and a simple functional form for utility. Other more flexible cases can be estimated similarly. For the purposes of this section, we assume that the data consists of a sample of consumers and includes their income, y_i, as well as their choice j in some market t.

For the purposes of this section, we will assume that the utility function takes the following form (omitting the t subscripts),

$$u_{ij} = log(x_j)\beta_{i,x} + log(\xi_j)\beta_{i,\xi} + log(y_i - p_j) \tag{31}$$

where $x_j \in \mathbb{R}^K$ and the coefficient on the $y_i - p_j$ term is normalized to 1 without loss of generality.

While we assume in this section that the researcher has access to micro data, it is not necessary in general to have micro data to use the techniques described in this paper. If only aggregate data is available, then the only difference in the approach would be that income effects would not be identified without further assumptions.[17]

[17] For example, it would be possible to identify income effects if the distribution of income were known and the distribution of income is assumed to be independent of preferences. This is the approach used in BLP. Alternatively, in many markets income effects are negligible and can be ignored.

Assuming an interior maximizer, equation (31) is maximized at

$$\frac{\beta_{i,k}}{x^*_{j,k}} = \frac{1}{y_i - p(x^*_j,\xi^*_j)}\frac{\partial p(x^*_j,\xi^*_j)}{\partial x_k} \quad \text{for } k \in 1..K \tag{32}$$

$$\frac{\beta_{i,\xi}}{\xi^*_j} = \frac{1}{y_i - p(x^*_j,\xi^*_j)}\frac{\partial p(x^*_j,\xi^*_j)}{\partial \xi} \tag{33}$$

where (x^*_j,ξ^*_j) represents the maximizing bundle. These first order conditions suggest the following estimator for the taste coefficients for individual i,

$$\hat{\beta}_{i,k} = \frac{x^i_{j,k}}{y_i - \hat{p}^i_j}\frac{\widehat{\partial p(x^i_j,\xi^i_j)}}{\partial x_k} \quad \text{for } k \in 1..K \tag{34}$$

$$\hat{\beta}_{i,\xi} = \frac{\hat{\xi}^i_j}{y_i - \hat{p}^i_j}\frac{\widehat{\partial p(x^i_j,\xi^i_j)}}{\partial \xi} \tag{35}$$

where $(x^i_j,\hat{\xi}^i_j)$ represents the (estimated) bundle chosen by individual i, $\hat{p}^i_j$ represents its estimated true price (with the measurement error removed), and $\frac{\widehat{\partial p(x^i_j,\xi^i_j)}}{\partial \xi}$ represents an estimator for the derivative of the price function at the chosen bundle.

Provided that an estimator is available for the derivatives of the price function (again see Matzkin (1999)), it is thus possible to estimate the vector of taste coefficients for each individual. Note that the asymptotic properties of the taste coefficient estimators depend only on the sample sizes for the first stage. Because of this, it is possible to obtain accurate estimates of the entire vector of taste coefficients for each individual using only a single observation.

Using the estimated taste coefficients for a sample of individuals, it is then possible to construct a density estimate of the distribution of taste coefficients in the population with only very weak additional assumptions. In particular, it is not necessary to place restrictions on the correlations between taste coefficients or the parametric form for the distribution of coefficients.

4.3 Estimation of Preferences, Discrete Case

In this section, we propose a simple Gibbs Sampling algorithm that can be used to estimate the distribution of consumer preferences in the case where the commodity space is discrete. To illustrate the approach, we consider a model where consumer i chooses from a finite set of $j = 1, ..., J$ products. Let consumer i's utility for product j satisfy:

$$u_{ij} = \beta_i \log(x_j) + \log(y_i - p_j) \tag{36}$$

where x_j and β_i are K-dimensional vectors and the coefficient on the outside good utility is normalized to one without loss of generality. Note that in this section we assume that the values of the unobserved characteristics have been estimated in a previous stage and are thus known. If consumer i chooses product j then it must be the case that product j maximizes consumer i's utility:

$$u_{ij}(\beta_i, x_j, y_i - p_j) \geq u_{ik}(\beta_i, x_k, y_i - p_k) \text{ for all } k \neq j \tag{37}$$

The likelihood function for this model is extremely simple. The probability that consumer i chooses product j is 1 if the set of inequalities in (37) are satisfied and 0 otherwise.[18] Let $x = (x_1, ..., x_J)$ be the vector of characteristics for products 1 through J and let $L(j|x, \beta_i)$ be the likelihood that consumer i chooses product j. Then

$$L(j|x, \beta_i) = \begin{cases} 1 & \text{if } u_{ij}(\beta_i, x_j, y_i - p_j) \geq u_{ik}(\beta_i, x_k, y_i - p_k) \text{ for all } k \neq j \\ 0 & \text{otherwise} \end{cases} \tag{38}$$

In Bayesian analysis, it is necessary to specify a prior distribution over parameters. This prior distribution will serve to identify the model subject to the identification discussion in section 3.8. That is, the prior allows us to choose one of the many possible aggregate taste distributions that are consistent with the choice data. The impact of the prior may be small or large depending on the characteristics space and the choice set. Its impact goes to zero as the number of products becomes large if the characteristics of the good are continuous. In that case the results from our Bayesian estimator would be consistent with classical econometric approaches. The Bayesian approach also allows the econometrician to incorporate auxiliary information that may lead to increased efficiency in the estimation. For instance, the econometrician can easily incorportate support restrictions or other restrictions on the β_i through the specification of the prior.

Many econometricians are hesitant to adopt a Bayesian approach to estimation because they feel that specifying a prior is inherently an ad hoc procedure. While it is true that a prior is based on the econometricians subjective beliefs it is worthwhile to keep three facts in mind. First, the problem we are studying is not identified. There is a set of parameters that will result in consumer i choosing product j. This problem can be handled in a straightforward fashion by adopting a Bayesian perspective. Second, it is straightforward to study the robustness of our results to alterations in the prior distribution of parameters. See Geweke (1997) for an overview of these procedures. Third, Bayesian analysis lends itself naturally to decision problems. For instance, Rossi (1997) uses the posterior distribution of parameters to study which individuals a profit maximimizing firm should coupon and what discount should be given in the coupon. For these reasons we believe that Bayesian methods are useful in the context of our particular problem. An alternative approach would be to calculate bounds on the aggregate distribution function.

For the purposes of our simple example, we will suppose that the prior distribution for β_i has a uniform distribution over the region B. Typically this region would be defined by a set of

[18]This abstracts from the possibility that a consumer has equal utility for two products. Since this is not a generic event we ignore this possibility.

conservative upper and lower bounds for each taste coefficient. We let $p(\beta_i)$ denote the prior distribution over the β_i. Let $C(i)$ denote the product chosen by household i. The posterior distribution for β_i, $p(\beta_i|C(i), x, p)$ conditional on the econometrician's information set then satisfies:

$$p(\beta_i|C(i), x, p) \propto \pi(\beta_i) L(j|x, \beta_i) \tag{39}$$

The posterior distribution will be a uniform distribution over those β_i that are consistent with the agents choice. So long as B completely covers all of the A_{ij} sets (see (24) for definition of A_{ij}), the posterior is uniform over A_{ij} for an individual i purchasing good j.

In applications, the econometrician is usually interested in some function of the parameter values $g(\beta_i)$ such as the posterior mean or the revenue a firm would receive from sending a coupon to send to household i. In our case we are interested in the value of the aggregate distribution function of the β_i's. We will cover estimation of that below. In general, the object of interest can be written as:

$$\int g(\beta_i) p(\beta_i|C(i), x, p) \tag{40}$$

One way to evaluate the above integral is by using Gibbs sampling. Gibbs sampling will generate a sequence of S pseudo-random parameters $\beta_i^{(1)}, \beta_i^{(2)}, ..., \beta_i^{(S)}$ with the property that:

$$\lim_{S \to \infty} \frac{1}{s} \sum_{s=1}^{S} g(\beta_i^{(s)}) = \int g(\beta_i) p(\beta_i|C(i), x, p) \tag{41}$$

In what follows, we will describe the mechanics of generating the set of pseudo-random parameters $\beta_i^{(1)}, \beta_i^{(2)}, ..., \beta_i^{(S)}$. Readers interested in a more detailed survey of Gibbs sampling can consult the surveys by Geweke (1994)or Geweke (1997).

Suppose that household i is observed to choose product j. The first step in developing a Gibbs sampler is to use equation (39) to find the following conditional distributions:

$$p(\beta_{i,1}|x, p, C(i) = j, \beta_{i,-1}) \tag{42}$$

$$p(\beta_{i,2}|x, p, C(i) = j, \beta_{i,-2}) \tag{43}$$

$$\vdots$$

$$p(\beta_{i,K}|x, p, C(i) = j, \beta_{i,-K}) \tag{44}$$

Since j is utility maximizing for household i it follows that:

$$\sum_l \beta_{i,l} \log(x_{l,j}) + \log(y_i - p_j) \geq \sum_l \beta_{i,l} \log(x_{l,k}) + \log(y_i - p_k) \text{ for all } k \neq j \tag{45}$$

which implies that:

$$\beta_{i,1} \geq \frac{\sum_{l\neq 1}\beta_{i,l}(\log(x_{l,k})-\log(x_{l,j}))+\log(y_i-p_j)-\log(y_i-p_k)}{\log(x_{1,j})-\log(x_{1,k})} \text{ if } x_{1,j} > x_{1,k} \quad (46)$$

$$\beta_{i,1} \leq \frac{\sum_{l\neq 1}\beta_{i,l}(\log(x_{l,k})-\log(x_{l,j}))+\log(y_i-p_j)-\log(y_i-p_k)}{\log(x_{1,j})-\log(x_{1,k})} \text{ if } x_{1,j} < x_{1,k} \quad (47)$$

$$(48)$$

Since the prior distribution is uniform and the likelihood is also uniform, it follows immediately that the the conditional distribution (42) must satisfy the inequalities implied by (46)-(47) and must also lie in the set B that defines the support of the prior.

To summarize, the conditional distribution (42) will be uniform on the interval $[\beta_{1,\min}, \beta_{1,\max}]$ where the support satisfies:

$$\beta_{1,\min} = \max\left\{\min_{\beta_1|\beta_{-1}} B, \max\left\{\frac{\sum_{l\neq 1}\beta_{i,l}(\log(x_{l,k})-\log(x_{l,j}))+\log(y_i-p_j)-\log(y_i-p_k)}{\log(x_{1,j})-\log(x_{1,k})} \text{ such that } x_{1,j} > x_{1,k}\right\}\right\} \quad (49)$$

$$\beta_{1,\max} = \min\left\{\max_{\beta_1|\beta_{-1}} B, \min\left\{\frac{\sum_{l\neq 1}\beta_{i,l}(\log(x_{l,k})-\log(x_{l,j}))+\log(y_i-p_j)-\log(y_i-p_k)}{\log(x_{1,j})-\log(x_{1,k})} \text{ such that } x_{1,j} < x_{1,k}\right\}\right\} \quad (50)$$

The conditional distribution for the remaining β's will also be a uniform distribution defined by inequalities that are analogous to (49) and (50).

Next, to evaluate the integral defined by (41) we need to generate a sequence $\beta_i^{(1)}, \beta_i^{(2)}, ..., \beta_i^{(S)}$ pseudo random numbers. Let $\beta_i^{(0)} = (\beta_{i,1}^{(0)}, \beta_{i,2}^{(0)})$ be an arbitrary point of support. We then use the following algorithm to generate the $\beta_i^{(1)}, \beta_i^{(2)}, ..., \beta_i^{(S)}$

1. Given $\beta_i^{(s)}$ draw $\beta_{i,1}^{(s+1)}$ from the distribution $p(\beta_{i,1}|x, p, C(i) = j, \beta_{i,-1}^{(s)})$.
2. Draw $\beta_{i,l}$ conditional on the vector $\beta_{i,-l}$ as in step 1, for $l = 2..K$.
3. Return to 1.

It can be easily verified that the sufficient conditions stated in Geweke (1994) are satisfied and that the simulation estimator defined in (41) will converge as the number of simulations tend to infinity.

The algorithm defined above is extremely simple to program since it merely requires the econometrician to draw a sequence of uniform random numbers. It is also straightforward to use Gibbs sampling to estimate far more general models of choice. All that is required is for the econometrician to find the conditional distributions similar to (42) and (43) for the given pure hedonic model and use Gibbs sampling.

In our case we are interested in recovering the distribution of tastes for the entire population of consumers. This involves a simple alteration of the algorithm above. Suppose that N_j out of a population of N consumers choose product j. Then the econometrician merely needs to simulate the posterior for consumers who choose product j and give each observation a weight of N_j/N.

It can easily be shown that if all of the characteristics are continuous, then the discrete model will converge to the continuous model at a rate of $\frac{1}{J}$ (see section 6.4). That is, as the number of products becomes large a consumer's true preference parameter will be perfectly learned, and in expectation the set that contains the consumer's preference parameter will have measure $\frac{1}{J}$. This is because the average measure of the A_{ij} sets is $\frac{1}{J}$.

In the above algorithm, we have assumed that all product characteristics are observed to the econometrician. This will not be true in general as we have emphasized in previous sections. One approach to this problem would to use an estimate of ξ_j obtained as in section 4.1.2 and proceed as above.[19]

A second difficulty that might be faced in practice is when there are no parameter values that satisfy the inequalities in equation (37) so that A_{ij} is empty for some i and j. Such an occurence is likely with strict functional forms for utility such as linear utility, particularly if the price function is also approximately linear. In cases such as this we would interpret that as a rejection of the functional form. Note that strict domination, where one product is "better" in all characteristics dimensions but has a lower price, is not possible in the two stage estimator because the first stage estimators of the unobserved characteristics rule this possibility out..

[19] In general, an estimate of ξ_j will have limited precision. It is possible to account for this in our estimation algorithm by using hierarchical Bayesian methods.

Suppose for instance that the prior distribution over the vector $\xi = (\xi_1, ..., \xi_J)$ is $p(\xi)$. Then given a random draw of ξ one can use the steps (1)-(3)in the above algorithm conditional upon a random draw of ξ from $p(\xi)$. Suppose that S_ξ such draws are made from $p(\xi)$. For each vector $\xi^{(s)}$, let $\beta_i^{(\xi^{(s)},1)}, \beta_i^{(\xi^{(s)},2)}, ..., \beta_i^{(\xi^{(s)},S)}$ be a vector of pseudo-random β_i's drawn from the above distribution.

Then it can easily be shown that:

$\lim_{S_\xi \to \infty} \lim_{S \to \infty} \frac{1}{SS_\xi} \sum_{s_\xi=1}^{S_\xi} \sum_{s=1}^{S} g(\beta_i^{(\xi^{(s_\xi)},s)}) = \int \int \{g(\beta_i) p(\beta_i | C(i), x, \xi, p) d\beta_i\} \ p(\xi) d\xi$

[Add calculation of the aggregate distribution function here.]

5 Conclusions

In this paper, we argued that current approaches to the estimation of random coefficient discrete choice models suffer from at least four problems. First, in many applications the instruments needed to identify the random coefficients may not be appealing. Second, it is not clear what variation in the data identifies the random coefficients. Third, commonly used models have unpleasant implications when the number of products becomes large. Fourth, it is computationally burdensome to estimate random coefficient models.

We demonstrated that random utility models build in a number of counter-intuitive properties as the number of products becomes large. We argued that in many applications, a pure hedonic model with an unobserved product characteristic is superior to random utility models. In general, the pure hedonic model with an unobserved product characteristic is not identified if the entire demand function is observed. However, if the unobserved product characteristic corresponds to a model or if the unobserved product characteristic is independent of the observed characteristic we show that the model is identified.

In many applications, the economist may not observe the entire demand function. Instead, the economist might only observe a finite number of choice per consumer. In this case, it is possible to recover the distribution of marginal rates of substitution at the chosen bundles. In most random coefficient models that are commonly used in empirical work, knowledge of the marginal utilities is sufficient to non-parametrically identify the population distribution of random coefficients (conditional on the functional form of utility).

In the case where the set of products is finite we developed a Gibbs sampling approach to simulate the posterior distribution of random coefficients. We demonstrated that if characteristics are continuous, then as the product space becomes sufficiently filled up, the Gibbs sampling algorithm will converge to an individual consumer's random coefficients and the population distribution of random coefficients can be recovered. The Gibbs procedure is also computationally quite simple.

6 Appendix

6.1 Proofs for Section 2

6.1.1 Proof of Property 1

Proof of Property 1. Consider demand at any point $(x_j, p_j, x_{-j}, p_{-j})$. Fix $\bar{\beta}_i \in B$ arbitrarily. If Ua holds, then choose $\bar{y}_i$ such that $\bar{y}_i > p_j$. Otherwise, choose $\bar{y}_i$ such that $f(\bar{y}_i) > 0$. Set $\epsilon_{ik} = 0$ for all $k \neq j$. Conditional on these values, product j is preferred to all other products if and only if

$$\epsilon_{ij} > \max_{k \neq j}\{u(x_k, \bar{y}_i - p_k, \bar{\beta}_i)\} - u(x_j, \bar{y}_i - p_j, \bar{\beta}_i) \equiv \bar{u}_k - \bar{u}_j. \tag{51}$$

By R(ii), the probability corresponding to (51) is strictly positive. Since y_i and β_i are independent of ϵ and $f_y(y_i) > 0$ and $f_\beta(\beta_i) > 0$,

$$q_j(x_j, p_j, x_{-j}, p_{-j}) > Prob[\epsilon_{ij} > \bar{u}_k - \bar{u}_j] f_y(y_i) f_\beta(\beta_i) > 0 \tag{52}$$

Since we chose price arbitrarily, this means that demand is positive for any price. □

6.1.2 Proof of Properties 2 and 5

Lemma 12. *Assumption R(ii) implies that* $\lim_{J\to\infty} \max_{j \in 1..J} \epsilon_i = \infty$ *a.s.*

Proof. For any $M < \infty$, let A_n be the event $\{\epsilon_{i0} < M, ..., \epsilon_{in} < M\}$. Then,

$$Pr(A_n) = Pr(\epsilon_{i0} < M) Pr(\epsilon_{i1} < M | \epsilon_{i0} < M) * ... * Pr(\epsilon_{in} < M | \epsilon_{i,-n} < M) \tag{53}$$

By assumption R2, there exists a δ_M such that each term in the above expression is less than δ_M. Therefore $Pr(A_n) < \delta_M^n$. Since this holds for all n, the sum $\sum_n Pr(A_n)$ must converge. By the Borell-Cantelli Lemma $Pr(\limsup A_n) = 0$. □

Properties 2, and 5 hold as a direct consequence of this.

6.1.3 Property 6

Proof of property 6. By the previous result,

$$\lim_{J\to\infty} \max_{j \in 0..J} u_{ij} = \infty \tag{54}$$

For any finite J, the compensating variation for removing the inside goods, CV, is the solution to,

$$u_0(y_i + \text{CV}) = (\max_{j \in 0..J} u_{ij}) - \epsilon_{i0} \tag{55}$$

For any given individual, the right hand side tends to infinity with J. Thus, it must be that CV does too. □

6.1.4 Properties 3 and 4

Proof of Properties 3 and 4. We show properties 3 and 4 for the *iid* case. This case provides the central intuition that the thickness of the tails of the distribution matters in determining the limiting properties of the demand system.

Assume that $\lim_{x\to\infty} \frac{f(x)}{1-F(x)} = \infty$ *(assumption H). Then: 1)* $\lim_{J\to\infty} E[\epsilon_1^J - \epsilon_2^J] = 0$*, where* ϵ_1^J *is the highest of* J *draws on* ϵ *and* ϵ_2^J *is the second highest, and 2) as the number of products becomes large the markup in a symmetric Bertrand-Nash price-setting equilibrium with single product firms tends to 0.*

1. Rewrite the desired expression using iterated expectations and bring the limit into the integral to get $E_{\epsilon_2^J}[lim_{J\to\infty} E(\epsilon_1^J - \epsilon_2^J \mid \epsilon_2^J)]$. Now, note that we have shown above that $lim_{J\to\infty}\epsilon_2^J = \infty$ a.s. It is also easy to show that $\lim_{x\to\infty} E[y - x \mid x] = 0$ if and only if the hazard rate of the conditional distribution $y|x$ goes to infinity as x becomes large. But, the conditional distribution of $\epsilon_1^J \mid \epsilon_2^J$ is proportional to the distribution of ϵ. Thus $\lim_{J\to\infty} E[\epsilon_1^J - \epsilon_2^J | \epsilon_2^J] = 0$ if and only if the hazard rate of $F()$ goes to infinity in the upper tail.

2. Consider J identical single product firms facing a demand system generated by a discrete choice model where the utility function is $u_{ij} = p_j - \epsilon_{ij}$ and the errors are assumed to be *iid*. In a symmetric Bertrand-Nash price setting equilibrium, all firms' prices will be the same and each firm will have equal market share $s_j = 1/J$. The markup will be $\frac{s_j}{-\frac{\partial s_j}{\partial p_j}}$. We now consider $\frac{\partial s_j}{\partial p_j}$:

$$s_j = Pr(\epsilon_k \leq \epsilon_j + p_k - p_j \forall k \neq j) \quad (56)$$

$$= \int_{-\infty}^{\infty} Pr(\epsilon_k \leq \epsilon_j + p_k - p_j \forall k \neq j \mid \epsilon_j) dP(\epsilon_j) \quad (57)$$

$$= \int_{-\infty}^{\infty} \Pi_{k \neq j} F(\epsilon_j + p_k - p_j) f(\epsilon_j) d\epsilon_j \quad (58)$$

$$= \int_{-\infty}^{\infty} F^{J-1}(\epsilon_j) f(\epsilon_j) d\epsilon_j \quad (59)$$

$$= 1/J \quad (60)$$

$$\Rightarrow \quad (61)$$

$$\frac{\partial s_j}{\partial p_j} = -\int_{-\infty}^{\infty} \sum_{k \neq j} \left(f(\epsilon_j + p_k - p_j) \Pi_{l \neq k,j} F(\epsilon_j + p_k - p_j) \right) f(\epsilon_j) d\epsilon_j \quad (62)$$

$$= -\int_{-\infty}^{\infty} (J-1) F^{J-2}(\epsilon_j) f(\epsilon_j) f(\epsilon_j) d\epsilon_j \quad (63)$$

$$\Rightarrow \quad (64)$$

$$1/\text{markup} = J \int_{-\infty}^{\infty} (J-1) F^{J-2}(\epsilon_j) f(\epsilon_j) f(\epsilon_j) d\epsilon_j \quad (65)$$

For the markup to go to zero the last expression must go to infinity. Note that $(J-1)F^{J-2}(\epsilon)f(\epsilon)$ is the density of ϵ_2^J so that the whole expression can be written as $E_{\epsilon_2^J}[Jf(\epsilon)]$. By markov's inequality, we have:

$$1/\text{markup} = E_{\epsilon_2^J}[Jf(\epsilon)] \quad (66)$$

$$\geq Jk_J Pr_{\epsilon_2^J}[Jf(\epsilon) \geq Jk_J] \quad (67)$$

$$= Jk_J Pr_{\epsilon_2^J}[f(\epsilon) \geq k_J] \quad (68)$$

$$= Jk_J \int_{-\infty}^{\infty} \{f(\epsilon) \geq k_J\} (J-1) F^{J-2}(\epsilon) f(\epsilon) d\epsilon \quad (69)$$

for any sequence k_J. This last expression makes it obvious that for any distribution whose density is bounded below (e.g. uniform) the markup does indeed converge to zero. We now show that this is also true for densities satisfying H.

Fix any $M < \infty$. Then by H there exists an $\underline{\epsilon}_M < \infty$ such that $\frac{f(\epsilon)}{1-F(\epsilon)} \geq M$ for all $\epsilon \geq \underline{\epsilon}_M$. Thus, for any number k_J, we have that if $\epsilon \geq \underline{\epsilon}_M$ and $M(1-F(\epsilon)) \geq k_J$ then it must be that $f(\epsilon) \geq k_J$. Now consider the integral above:

$$Jk_J \int_{-\infty}^{\infty} \{f(\epsilon) \geq k_J\} (J-1) F^{J-2}(\epsilon) f(\epsilon) d\epsilon \geq Jk_J \int_{\underline{\epsilon}_M}^{\infty} \{M(1-F(\epsilon)) \geq k_J\} (J-1) F^{J-2}(\epsilon) f(\epsilon) d\epsilon \quad (70)$$

However, we can now solve for the upper end of the region of integration as well because $F()$

is a monotonic function:

$$M(1-F(\epsilon)) \geq k_J \tag{71}$$

$$\Leftrightarrow F(\epsilon) \leq 1-\frac{k_J}{M} \tag{72}$$

$$\Leftrightarrow \epsilon \leq F^{-1}(1-\frac{k_J}{M}) \tag{73}$$

Plugging this back into the integral gives:

$$Jk_J \int_{-\infty}^{\infty} \{f(\epsilon) \geq k_J\}(J-1)F^{J-2}(\epsilon)f(\epsilon)d\epsilon \geq Jk_J \int_{\underline{\epsilon}_M}^{F^{-1}(1-\frac{k_J}{M})} (J-1)F^{J-2}(\epsilon)f(\epsilon)d\epsilon \tag{74}$$

$$= Jk_J \left[(1-\frac{k_J}{M})^{J-1} - F^{J-1}(\underline{\epsilon}_M)\right] \tag{75}$$

$$= Jk_J \left(1-\frac{k_J}{M}\right)^{J-1} - Jk_J\delta^{J-1} \tag{76}$$

where $\delta = F(\underline{\epsilon}_M) < 1$. We now let $k_J = J^{-1/\gamma}$ where $\gamma > 1$. The second part of the expression goes to zero as J gets large (since the exponential portion goes to zero faster than J). The rate of convergence of k_J has been chosen such that the first part diverges. □

6.1.5 Property 7

Proof of Property 7. Loose Sketch of Proof (Note: we have not proven this satisfactorily yet but we believe it to be true.) Suppose that the error term ϵ_{ij} represented unobserved individual tastes and unobserved product characteristics. Denote the unobserved vector of individual tastes as λ_i and the unobserved vector of product characteristics as η_j and let $\epsilon_{ij} = f(\lambda_i, \eta_j)$, for some function $f(\cdot)$. Then the dimension of ϵ would be at most equal to the dimension of λ plus the dimension of η. Suppose that the dimension of λ were fixed at a and the dimension of η were fixed at b. Then the dimension of ϵ would be at most $a*I+b*J$. Thus, with a fixed number of unobservable individual and product characteristics, it is not in general possible to generate the error term in standard discrete choice econometric models, since that error term has dimension $I*J$. Therefore, the interpretation of the error term in those models as unobservable individual and product characteristics is not correct. There are only two ways to generate the error properties of the standard econometric model using unobservable tastes and product characteristics. Either the number of unobservable tastes, a, must grow with the number of products, or the number of unobservable product characteristics, b, must grow with the number of individuals, or both. In either case, the error term can be interpreted as a "taste for products". □ □

6.2 Proof of Theorem 1

Proof. (i) Suppose $p_{jt} > p_{j't}$ for some t. Then since u_i is strictly increasing in c, $u_i(x_j, \xi_j, y_i - p_{jt}) < u_i(x_{j'}, \xi_{j'}, y_i - p_{j't})$ for all individuals. This implies that demand for j is zero in market t, which is a contradiction.

(ii) Suppose $p_{jt} \leq p_{j't}$ for some t. Then since u_i is strictly increasing in c and strictly increasing in ξ, $u_i(x_j, \xi_j, y_i - p_{jt}) > u_i(x_{j'}, \xi_{j'}, y_i - p_{j't})$ for all individuals. This implies that demand for j' is zero in market t, which is a contradiction.

(iii) If j and j' have the same prices, then the result holds trivially. This also covers the case where j and j' have the same characteristics because of (i). Suppose that j and j' have different characteristics in at least one dimension and assume without loss of generality that $p_{jt} > p_{j't}$. Since u_i is Lipschitz continuous in (x_j, ξ_j), we have that

$$|u_i(x_j, \xi_j, y_i - p_{jt}) - u_i(x_{j'}, \xi_{j'}, y_i - p_{jt})| \leq M_1(|x_j - x_{j'}| + |\xi_j - \xi_{j'}|) \tag{77}$$

By a mean value expansion, for all individuals,

$$u_i(x_{j'}, \xi_{j'}, y_i - p_{j't}) = u_i(x_{j'}, \xi_{j'}, y_i - p_{jt}) + (p_{jt} - p_{j't})\frac{\partial u_i(x_{j'}, \xi_{j'}, y_i - p^*_{jt})}{\partial c} \tag{78}$$

where $p^*_{jt} \in [p_{jt}, p_{j't}]$ and varies for each i. Plugging (78) into (77) gives

$$\begin{aligned} \Big|(u_i(x_j, \xi_j, y_i - p_{jt}) - u_i(x_{j'}, \xi_{j'}, y_i - p_{j't})) + (p_{jt} - p_{j't})\frac{\partial u_i(x_{j'}, \xi_{j'}, y_i - p^*_{jt})}{\partial c}\Big| \\ \leq M_1(|x_j - x_{j'}| + |\xi_j - \xi_{j'}|) \end{aligned} \tag{79}$$

The second term in the absolute value on the left hand side is positive. Since demand for j is positive, there must be some individuals for which the first term is also positive. For those individuals, we can ignore the absolute value sign and we only strengthen the inequality by also ignoring the first term,

$$\begin{aligned}
(p_{jt} - p_{j't})\frac{\partial u_i(x_{j'}, \xi_{j'}, y_i - p^*_{jt})}{\partial c} &\leq M_1(|x_j - x_{j'}| + |\xi_j - \xi_{j'}|) \text{ for } i \text{ that prefer } j \text{ to } j'. && (80) \\
(p_{jt} - p_{j't}) &\leq \frac{M}{\frac{\partial u_i(x_{j'}, \xi_{j'}, y_i - p^*_{jt})}{\partial c}}(|x_j - x_{j'}| + |\xi_j - \xi_{j'}|) \text{ for } i \text{ that prefer } j \text{ to } j'. && (81) \\
&\leq \frac{M_1}{\epsilon}(|x_j - x_{j'}| + |\xi_j - \xi_{j'}|) && (82) \\
&= M_2(|x_j - x_{j'}| + |\xi_j - \xi_{j'}|) && (83)
\end{aligned}$$

□

6.3 Proofs for section 4.1.1

Some definitions:

Definition 7. *Let $\mathcal{K}_{k,p}$ be the class of bounded Borel measurable real-valued functions $K(\cdot)$ on $\mathbb{R}^K$ such that, with $z = (z_1, ..., z_K)'$, $z_i \in \mathbb{R}$,*

$$\int z_1^{i_1} z_2^{i_2} \cdots z_K^{i_K} K(z) dz_1 \cdots dz_K = \begin{cases} 1 & \text{if } i_1 = i_2 = \cdots = i_K = 0 \\ 0 & \text{if } 0 < i_1 + i_2 + \cdots + i_k < p \end{cases}$$

$$\int |z|^i |K(z)| dz < \infty \text{ for } i = 0 \text{ and } i = p$$

$$\int K(z) dz = 1$$

Definition 8. *Let $\mathcal{D}_{K,p}$ be the class of all continuous real-valued functions f on $\mathbb{R}^K$ such that the derivatives*

$$\frac{\partial^I f(z)}{\partial^{i_1} z_1 \partial^{i_2} z_2 \cdots \partial^{i_K} z_K} \quad I \equiv \sum_{j=1}^{K} i_j, \; i_j \geq 0$$

are continuous and uniformly bounded for $0 \leq I \leq p$.

6.3.1 Lemma for $\bar{\epsilon}$

Let $\bar{\epsilon}_k = \frac{1}{T} \sum_{t=1}^{T} \epsilon_{kt}$. We will assume that ϵ_{kt} is *iid*, mean zero, with finite variance,

Assumptions:

E1 ϵ_{kt} is *iid* with $E(\epsilon_{kt}) = 0$, $Var(\epsilon_{kt}) = \sigma_k^2$ for all k, and $E|\epsilon_{kt}|^r$ exists for some $2 \leq r \leq \infty$.

E2 $J^{-2/r} T h_J^2 \to \infty$, where r is as in E1.

Lemma 13. *Under E1-E2, $sup_k |\frac{\bar{\epsilon}_k}{h_J}| \to 0$ in probability.*

Proof. Without loss of generality, we order the ϵ's such that $k = 1$ refers to the ϵ with the highest variance σ_k^2.

$$
\begin{aligned}
Pr(sup_k|\frac{\bar{\epsilon}_k}{h_J}| < \delta) &= Pr(sup_k|\bar{\epsilon}_k| < \delta h_J) &(84)\\
&= \Pi_{k=1}^J Pr(|\bar{\epsilon}_k| \leq \delta h_J) &(85)\\
&\geq Pr(|\bar{\epsilon}_1| \leq \delta h_J)^J &(86)\\
&\geq \left(1 - \frac{E|\bar{\epsilon}_1|^r}{\delta^r h_J^r}\right)^J &(87)\\
&= ((1 - z_J)^{1/z_J})^{z_J J} &(88)
\end{aligned}
$$

where the first inequality holds by the ordering of the variances, the second holds by Chebyshev's inequality, and

$$
\begin{aligned}
z_J &= \frac{E|\bar{\epsilon}_1|^r}{\delta^2 h_J^r}\\
&= \frac{T^{r/2} E|\bar{\epsilon}_1|^r}{T^{r/2}\delta^2 h_J^r}\\
&= \frac{E|T^{1/2}\bar{\epsilon}_1|^r}{(T^{1/2}\delta h_J)^r}
\end{aligned}
$$

The result of the lemma holds by (88) if $z_J J = o(1)$ (because the term inside the first bracket tends to e^{-1}). By a CLT, the numerator of z_J is $O(1)$. J^{-1} times the denominator of z_J diverges by E2. □

6.3.2 Proofs for $\hat{g}$

Let,

$$
\hat{g}_t^T(x_0, \bar{p}_0^T) = \frac{\frac{1}{J}\sum_{k=1}^J \frac{1}{h_J^{K+1}} K_1\left(\frac{\bar{p}_0^T - \bar{y}_k^T}{h_J}\right) K_2\left(\frac{x_0 - x_k}{h_J}\right) y_{kt}}{\frac{1}{J}\sum_{k=1}^J \frac{1}{h_J^{K+1}} K_1\left(\frac{\bar{p}_0^T - \bar{y}_k^T}{h_J}\right) K_2\left(\frac{x_0 - x_k}{h_J}\right)} \quad (89)
$$

and let,

$$
\hat{\hat{g}}_t^T(x_0, \bar{p}_0^T) = \frac{\frac{1}{J}\sum_{k=1}^J \frac{1}{h_J^{K+1}} K_1\left(\frac{\bar{p}_0^T - \bar{p}_k^T}{h_J}\right) K_2\left(\frac{x_0 - x_k}{h_J}\right) y_{kt}}{\frac{1}{J}\sum_{j=1}^J \frac{1}{h_J^{K+1}} K_1\left(\frac{\bar{p}_0^T - \bar{p}_k^T}{h_J}\right) K_2\left(\frac{x_0 - x_k}{h_J}\right)} \quad (90)
$$

Then

$$\hat{g}_t^T(x_j,\bar{y}_j^T) - p_t(x_j,\xi_j) = (\hat{\hat{g}}_t^T(x_j,\bar{p}_j^T) - p_t(x_j,\xi_j)) + (\hat{g}_t^T(x_j,\bar{y}_j^T) - \hat{\hat{g}}_t^T(x_j,\bar{p}_j^T)) \tag{91}$$

Uniform Consistency:

The first term of equation (91) is standard. Thus,

$$\sup_{\{(x_j,\xi_j)\in\mathbb{R}^{K+1}:h(x_j,\xi_j)>\delta\}} |\hat{\hat{g}}_t^T(x_j,\bar{p}_j^T) - p_t(x_j,\xi_j)| \to 0$$

in probability.

That leaves the second term. Consider the numerator of the second term first,

$$\begin{aligned}
&(\hat{g}_t^T(x_j,\bar{y}_j^T)\hat{h}(x_j,\bar{y}_j^T) - \hat{\hat{g}}_t^T(x_j,\bar{p}_j^T)\hat{\hat{h}}(x_j,\bar{p}_j^T)) \\
&= \frac{1}{Jh_J^{K+1}}\sum_{k=1}^{J}\left[K_1\left(\frac{\bar{y}_j^T-\bar{y}_k^T}{h_J}\right) - K_1\left(\frac{\bar{p}_j^T-\bar{p}_k^T}{h_J}\right)\right]K_2\left(\frac{x_j-x_k}{h_J}\right)y_{kt}
\end{aligned} \tag{92}$$

$$\begin{aligned}
&= \frac{1}{Jh_J^{K+1}}\sum_{k=1}^{J}\left(\frac{\bar{\epsilon}_j^T-\bar{\epsilon}_k^T}{h_J}\right)K_1'\left(\frac{\bar{p}_j^T-\bar{p}_k^T}{h_J}\right)K_2\left(\frac{x_j-x_k}{h_J}\right)y_{kt} \\
&\quad + \frac{1}{2}\frac{1}{Jh_J^{K+1}}\sum_{k=1}^{J}\left(\frac{\bar{\epsilon}_j^T-\bar{\epsilon}_k^T}{h_J}\right)^2 K_1''\left(\lambda_{jk}^T\right)K_2\left(\frac{x_j-x_k}{h_J}\right)y_{kt}
\end{aligned} \tag{93}$$

where $\lambda_{jk}^T \in \left[\frac{\bar{p}_j^T-\bar{p}_k^T}{h_J}, \frac{\bar{p}_j^T-\bar{p}_k^T}{h_J} + \frac{\bar{\epsilon}_j^T-\bar{\epsilon}_k^T}{h_J}\right]$ and $\lambda_{jk}^T \to \frac{\bar{p}_j^T-\bar{p}_k^T}{h_J}$ in probability.

Note that

$$\frac{\bar{\epsilon}_j^T-\bar{\epsilon}_k^T}{h_J} = \frac{\bar{\epsilon}_j^T-\bar{\epsilon}_{kt}^{T-1}}{h_J} - \frac{\epsilon_{kt}}{Th_j},$$

where $\bar{\epsilon}_{kt}^{T-1} = \frac{1}{T}\sum_{s\neq t}\epsilon_{ks}$. Substituting in to (93) gives,

$$
\begin{aligned}
&= \frac{1}{Jh_J^{K+1}}\sum_{k=1}^{J}\frac{\bar{\epsilon}_j^T - \bar{\epsilon}_{kt}^{T-1}}{h_J}K_1'\left(\frac{\bar{p}_j^T - \bar{p}_k^T}{h_J}\right)K_2\left(\frac{x_j - x_k}{h_J}\right)y_{kt} \\
&- (Th_J)^{-1}\frac{1}{Jh_J^{K+1}}\sum_{k=1}^{J}K_1'\left(\frac{\bar{p}_j^T - \bar{p}_k^T}{h_J}\right)K_2\left(\frac{x_j - x_k}{h_J}\right)y_{kt}\epsilon_{kt} \\
&+ \frac{1}{2Jh_J^{K+1}}\sum_{k=1}^{J}\frac{(\bar{\epsilon}_j^T - \bar{\epsilon}_{kt}^{T-1})^2}{h_J^2}K_1''(\lambda_{jk}^T)K_2\left(\frac{x_j - x_k}{h_J}\right)y_{kt} \\
&- (Th_J^2)^{-1}\frac{1}{Jh_J^{K}}\sum_{k=1}^{J}\frac{\bar{\epsilon}_j^T - \bar{\epsilon}_{kt}^{T-1}}{h_J}K_1''(\lambda_{jk}^T)K_2\left(\frac{x_j - x_k}{h_J}\right)y_{kt}\epsilon_{kt} \\
&+ (Th_J)^{-2}\frac{1}{2Jh_J^{K}}\sum_{k=1}^{J}K_1''(\lambda_{jk}^T)K_2\left(\frac{x_j - x_k}{h_J}\right)y_{kt}\epsilon_{kt}^2
\end{aligned}
\tag{94}
$$

$$
\begin{aligned}
&\leq \sup_k\left|\frac{\bar{\epsilon}_j^T - \bar{\epsilon}_{kt}^{T-1}}{h_J}\right|\frac{1}{Jh_J^{K+1}}\sum_{k=1}^{J}\left|K_1'\left(\frac{\bar{p}_j^T - \bar{p}_k^T}{h_J}\right)K_2\left(\frac{x_j - x_k}{h_J}\right)\right||y_{kt}| \\
&+ (Th_J)^{-1}\left|\frac{1}{Jh_J^{K+1}}\sum_{k=1}^{J}K_1'\left(\frac{\bar{p}_j^T - \bar{p}_k^T}{h_J}\right)K_2\left(\frac{x_j - x_k}{h_J}\right)y_{kt}\epsilon_{kt}\right| \\
&+ \sup_k\frac{(\bar{\epsilon}_j^T - \bar{\epsilon}_{kt}^{T-1})^2}{2h_J^3}sup_\lambda|K_1''(\lambda)|\frac{1}{Jh_J^K}\sum_{k=1}^{J}\left|K_2\left(\frac{x_j - x_k}{h_J}\right)\right||y_{kt}| \\
&+ (Th_J^2)^{-1}\sup_k\left|\frac{\bar{\epsilon}_j^T - \bar{\epsilon}_{kt}^{T-1}}{h_J}\right|\sup_\lambda|K_1''(\lambda)|\frac{1}{Jh_J^K}\sum_{k=1}^{J}\left|K_2\left(\frac{x_j - x_k}{h_J}\right)\right||y_{kt}\epsilon_{kt}| \\
&+ (Th_J)^{-2}\sup_\lambda|K_1''(\lambda)|\frac{1}{2Jh_J^K}\sum_{k=1}^{J}\left|K_2\left(\frac{x_j - x_k}{h_J}\right)\right||y_{kt}\epsilon_{kt}^2|
\end{aligned}
\tag{95}
$$

The second and fifth terms converge in probability to zero uniformly over (x_j, ξ_j) by standard results. The first and fourth terms converge in probability to zero uniformly over (x_j, ξ_j) by standard results and the lemma above. The third term converges more slowly in T than any of the others due to the extra h_J term in the denominator. Using only a second order expansion, in order for this term to converge to zero it is necessary that $J^{-r/2}Th_J^3 \to \infty$ (by the lemma above). However, using a higher order expansion, the required convergence rate for T can be slowed to that listed in C5.

The denominator of the second term in (91) can be treated similarly by changing all of the y_{kt} terms above to 1's. Thus, uniform consistency of the whole second term is obtained on a set where $h(x) > \delta$ for some $\delta > 0$.

Asymptotic Normality

By standard results, the asymptotic distribution of the first term is,

$$\sqrt{Jh_J^{K+1}}[\hat{\hat{g}}_t^T(x_j, \bar{p}_j^T(x_j, \xi_j)) - p_t(x_j, \xi_j)] \to N\left(\frac{\lambda b(x_j, \xi_j)}{f(x_j, \xi_j)}, \ \frac{\sigma_\epsilon^2(x_j, \xi_j)}{f(x_j, \xi_j)} \int K(r)^2 dr\right)$$

where

$$\begin{aligned} & b(x_j, \xi_j) \\ &= \lim_{J \to \infty} E[g(x_i, \bar{p}^T(x_i, \xi_i)) - g(x_j, \bar{p}^T(x_j, \xi_j))] K_1\left(\frac{\bar{p}^T(x_j, \xi_j) - \bar{p}^T(x_i, \xi_i)}{h_J}\right) K_2\left(\frac{x_j - x_i}{h_J}\right) h_J^{-m-(K+1)} \end{aligned}$$

and

$$\sigma_\epsilon^2(x_j, \xi_j) = E(\epsilon_j^2 | x = x_j, \xi = \xi_j).$$

To show the result, we again rely on the breakdown in (91) and the bound for the second term provided by (95). By the lemma above and standard results is easy to show that under assumptions C5 and C6 the five terms in (95) converge to zero faster than $\sqrt{J}h_J^{K+1}$. Therefore the fact that the estimated average prices are used in place of the actual average prices does not affect the asymptotic distribution of the estimator.

6.4 Convergence of the Discrete Model to the Continuous Model

In this section, we discuss the problem of identification for the discrete model. Clearly, when the number of products is finite, it will no longer be the case that person i's coefficients β_i can be uniquely recovered. Often, it will be possible to make a small perturbation to the β_i and leave choice behavior unaltered. A first result we establish is that as the number of products becomes infinite, then the β_i will be uniquely determined. The continuous model is therefore a limiting case of the discrete model. In practice, this observation yields powerful insights into the identification of discrete choice models. As we demonstrated in the previous section, with a continuum of products choice behavior and identification follow in a straightforward fashion from standard microeconomic theory. When the set of products is sufficiently large these simple insights can be used to study the problem of identification in the discrete case as well.

We begin by considering the case where all product characteristics are observable to both the consumer and the econometrician. We write consumer i's utility as $u_{ij} = u(x_j, p_j, \beta_i)$. Also, suppose that there is a pricing function $p(x)$ that maps characteristics into prices in the sense that $p_j = p(x_j)$ for any product j. We now make two assumptions about the product space and the utility:

Assumption 1. All of the product characteristics x_j are elements of x an open, bounded and convex subset of R^N. Also, all of the β_i lie in B, an open, bounded and convex subset of R^N.

Assumption 2. For any β_i, the function $u(x, p(x), \beta_i)$ is strictly concave and continuously differentiable. Furthermore, the matrix $D_{\beta,x}u(x, p(x), \beta_i)$ has full rank for all x and for all β.

Assumption 3. The dimension of the vector β_i is at least as great as x. Also, suppose that for every element of $x \in x$ there exists a β_i for which x is a utiltiy maximizing choice in x.

Next, we make precise the sense in which the discrete model is a limit of the continuous model. Suppose that we draw a random sequence $x^{(1)}, x^{(2)}, ..., x^{(n)}, ...$ of products from x. Let $S^{(n)} = \{x^{(1)}, x^{(2)}, ..., x^{(n)}\}$ be a set of choices available to consumer i that is comprised of the first n elements of our sequence. Let $C(n)$ be the utility maximizing choice for consumer i when she can choose from $S^{(n)}$.

Theorem 14. *Let $B^{(n)}$ denote those elements of B such that $C(n)$ is the utility maximizing choice in the set $S^{(n)}$. Suppose that Assumptions 1-3 hold. Then with probability one, $\lim_{n\to\infty} B^{(n)} = \beta_i$.*

Proof. Let x^* be the utility maximizing product for a household with random coefficients β_i when the entire set of products x is available. We start by demonstrating that as $n \to \infty$ it must be the case that $\lim_{n\to\infty} C(n)$ exists and is equal to x^*. By the definition of stict concavity, the set of products $U^{(n)}$ that give utility strictly greater than $C(n)$ is open and strictly convex. Therefore, the sequence of sets $U^{(n)}$ is decreasing in the sense that $U^{(n)} \supseteq U^{(n+1)}$. Let $U^* = \cap_{n=1}^{\infty} U^{(n)}$. It must be the case that $x^* \in U^*$ since x^* maximizes utility on the entire set x. Furthermore, $x^* = U^*$. To see why, suppose that there is another element $x' \in U^*$. By the continuity of utility, it is possible to find an open neighborhood of x^* every element of which gives strictly higher utility than x'. Since this is an open neighborhood with probability one some $x^{(n)}$ will be in this neighborhood and therefore $x' \notin U^{(n)}$ which is a contradiction. Finally, since $D_{\beta,x}u(x, p(x), \beta_i)$ has full rank, it is possible to invert the first order conditions as in the proof of Theorem 1 to identify β_i from knowledge of x^*. Let $B^* = \cap B^{(n)} \to \beta_i$. The set U^* must contain at least as many elements as the set B^*. For each element of $\beta_i' \in B^*$ it must be the case that $argmax_{x \in x}\ u(x, p(x), \beta_i)$ is in U^*. By assumption 3, for each β_i' that value of $argmax_{x \in x}\ u(x, p(x), \beta_i)$ is unique. Therefore, B^* is a singleton. ∎

Note that it would be possible to extend our proof to cover the case where there is at least one unobserved product characteristic. The proof would have to be modified so that it was possible to retreive the unobserved product characteristic by a hedonic regression as in the previous sections. This is entirely straightforward, but very tedious so we omit this extension.

7 References

Ackerberg, D. and M. Rysman (2000), "Unobserved Product Differentiation in Discrete Choice Models: Estimating Price Elasticities and Welfare Effects," Working Paper, Boston University.

Anderson, S., de Palma, A. and J. Thisse (1992), *Discrete Choice Theory of Product Differentiation,* Cambridge: MIT Press.

Bajari, P. and M. Kahn (2000), "Why Does Racial Segregation Persist? A Case Study of Philadelphia," Working Paper, Stanford University.

Benkard, C. (2000), "A Dynamic Analysis of the Market for Wide-Bodied Commercial Aircraft," NBER Working Paper, #W7710.

Berry, S. (1994), "Estimating Discrete-Choice Models of Product Differentiation," *RAND Journal of Economics*, **25:2**, 242-262.

Berry, S., J. Levinsohn, and A. Pakes (1995), "Automobile Prices in Market Equilibrium," *Econometrica*, **63:4**, 841-89.

Berry, S., J. Levinsohn, and A. Pakes (1998), "Differentiated Product Demand Systems From a Combination of Micro and Macro Data: The New Car Market," NBER Discussion Paper, #6481.

Berry, S., and A. Pakes (2000), "Estimating the Pure Hedonic Discrete Choice Model," Working Paper, Yale University.

Blundell, R. and Powell (2000), "Endogeneity in Nonparametric and Semiparametric Regression Models," Working Paper, Econometric Society World Meetings, Seattle, August 2000.

Bresnahan, T. and R. Gordon (1997), *The Economics of New Goods*, Studies in Income and Wealth, Vol. 58, Chicago: NBER.

Caplin, A. and B. Nalebuff (1991), "Aggregation and Imperfect Competition: On the Existence of Equilibrium," *Econometrica*, **59:1**, 1-23.

Davis, P. (2000) "Spatial Competition in Retail Markets: Movie Theatres," Working Paper, MIT Sloan.

Epple, D. (1987), "Hedonic Prices and Implicit Markets: Estimating Demand and Supply Functions for Differentiated Products," *Journal of Political Economy*, **95:1**, 59-80.

Feenstra, R. (1995), "Exact Hedonic Price Indexes," *Review of Economics and Statistics*, 634-653.

Goettler, R. and R. Shachar (1999), "Estimating Product Characteristics and Spatial Competition in the Network Television Industry," Working Paper, Carnegie Mellon University.

Hausman, J. (1997), "Valuation of New Goods Under Perfect and Imperfect Competition," in Bresnahan, T. and R. Gordon (eds.), *The Economics of New Goods*, Studies in Income and Wealth, Vol. 58, Chicago: NBER.

Hendel, I. (1999), "Estimating Multiple-Discrete Choice Models: An Application to Computerization Returns," *Review of Economic Studies*, **66:2**, 423-446.

Hurwicz, L. and H. Uzawa (1971), "On the Integrability of Demand Functions," Chapter 6 in *Preferences, Utility, and Demand,* ed. By John S. Chipman, Leonid Hurwicz, Marcel K. Richter, and Hugo F. Sonnenschein, New York: Harcourt Brace Jovanovich, Inc., 1971.

Jones, L. (1984), "A Competitive Model of Commodity Differentiation," *Econometrica*, **52:2**, March 1984.

Mas-Colell, A. (1975), "A Model of Equilibrium with Differentiated Commodities," *Journal of Mathematical Economics*, **2:2**, 263-295.

Matzkin, Rosa (1999), "Nonparametric Estimation of Nonadditive Random Functions," Mimeo, Northwestern University.

Nevo, A. (2000), "Measuring Market Power in the Ready-To-Eat Cereal Industry," *Econometrica*, forthcoming.

Petrin, A. (1998), "Quantifying the Benefits of New Products: The Case of the Minivan," Working Paper, Graduate School of Business, University of Chicago.

Rosen, S. (1974), "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *Journal of Political Economy*, **82**, 34-55.

Trajtenburg, M. (1989), "The Welfare Analysis of Product Innovations: with an Application to Computed Tomography Scanners," *Journal of Political Economy*, **97**, 444-479.