TECHNICAL WORKING PAPER SERIES

DEMAND ESTIMATION WITH HETEROGENEOUS CONSUMERS AND UNOBSERVED PRODUCT CHARACTERISTICS: A HEDONIC APPROACH

Patrick Bajari
C. Lanier Benkard

Technical Working Paper 272
http://www.nber.org/papers/T0272

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2001

We would like to thank Daniel Ackerberg, Steven Berry, Richard Blundell, Timothy Bresnahan, Donald Brown, Ian Crawford, Hidehiko Ichimura, Guido Imbens, John Krainer, Jonathon Levin, Rosa Matzkin, Costas Meghir, Whitney Newey, Ariel Pakes, Peter Reiss, Marcel Richter, and Ed Vytlacil for many helpful discussions, as well as seminar participants at Carnegie Mellon, Northwestern, Stanford, UBC, UCL, UCLA, Wash. Univ. in St. Louis, Wisconsin, and Yale for helping us to clarify our thoughts. Both authors would also like to thank the Hoover Institution for their present and future support. Any remaining errors are our own. The views expressed in this paper are those of the authors and not necessarily those of the National Bureau of Economic Research.

Demand Estimation With Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach
Patrick Bajari and C. Lanier Benkard
NBER Technical Working Paper No. 272
July 2001
JEL No. D1, C1

ABSTRACT

We study the identification and estimation of preferences in hedonic discrete choice models of demand for differentiated products. In the hedonic discrete choice model, products are represented as a finite dimensional bundle of characteristics, and consumers maximize utility subject to a budget constraint. Our hedonic model also incorporates product characteristics that are observed by consumers but not by the economist. We demonstrate that, unlike the case where all product characteristics are observed, it is not in general possible to uniquely recover consumer preferences from data on a consumer's choices. However, we provide several sets of assumptions under which preferences can be recovered uniquely, that we think may be satisfied in many applications. Our identification and estimation strategy is a two stage approach in the spirit of Rosen (1974). In the first stage, we show under some weak conditions that price data can be used to nonparametrically recover the unobserved product characteristics and the hedonic pricing function. In the second stage, we show under some weak conditions that if the product space is continuous and the functional form of utility is known, then there exists an inversion between a consumer's choices and her preference parameters. If the product space is discrete, we propose a Gibbs sampling algorithm to simulate the population distribution of consumers' taste coefficients.

Patrick Bajari
Department of Economics
Stanford University
Stanford, CA 94305
and NBER

C. Lanier Benkard
Graduate School of Business
Stanford University
Stanford, CA 94305-5015
and NBER

1 Introduction

In this paper we study the problem of identification and estimation of preferences in hedonic discrete choice models of demand for differentiated products. This class of models includes Rosen's (1974) model, as well as standard econometric models such as logit and probit, and random coefficients versions of these models. 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 preferences are 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. For example, using revealed preference arguments applied at the individual level, it is often possible to learn about the shape of product demand curves in areas of price space that are not observed in the data.

Like much of the recent empirical literature on demand estimation (e.g., Ackerberg and Rysman (2000), Berry (1994), Berry and Pakes (2000), Berry, Levinsohn, and Pakes (1995) [BLP], Berry, Levinsohn and Pakes (1998), Davis (2000), Goettler and Shachar (1999), Hendel (1999), Nevo (2000), Petrin (1998), McCulloch and Rossi (1996), and others), we focus on a model in which products are defined as vectors of characteristics. However, one of our primary goals is to investigate to what extent it is possible to identify consumers' utility functions over characteristics in general. Thus, in this paper we begin by considering a more general model of preferences than has commonly been applied, in which the functional form of utility and the distributions of taste coefficients are not necessarily known.[1]

[1] However, we note here that one of our conclusions will be that if the data contains only a single choice observation per individual, then it will not be possible to identify each individual's preferences without functional form assumptions similar to those used in the previous literature.

The characteristics approach to modeling demand for differentiated products has also been used extensively in the past. Examples include models of horizontal product differentiation such as Hotelling (1929), Gorman (1980), and Lancaster (1966), models of vertical product differentiation such as Shaked and Sutton (1987) and Bresnahan (1987), as well as Rosen's (1974) model which is the basis of the two-stage hedonics approach.

One thing that differentiates this paper from the past literature on hedonic models is that we study the identification of preferences when one product characteristic is not observed by the economist. In section 2.3, we establish that, even with many observations on a consumer's choices (such that the consumer's entire demand function is known), it is not in general possible to uniquely recover consumer preferences. The intuition for this result, which is is an extension of the main result of Varian (1988),[2] is that in a wide class of models it is possible to attribute all of the consumer's utility to the unobserved product characteristics. This result stands in sharp contrast to the uniqueness results of Mas-Collel (1977), which suggest that preferences can be recovered uniquely from observed choices if all product characteristics are observed.

In section 4 we show that individual preferences are only just identified if all product characteristics are observed. An important implication of this result is that, when estimating preferences, the choice data contains no additional information that can be used to estimate the unobserved product characteristics. We instead propose a strategy for identifying the unobserved characteristics using information contained in prices. The intuition for our identification strategy is that if two products have the same observed characteristics but one has a higher price, then it must be that this product is better in the unobservable dimension, or otherwise it would not have positive demand.

Thus, our identification and estimation strategy has two stages, in the spirit of Rosen (1974). In the first stage, the price function and product unobservables are estimated using data on prices and characteristics. In the second stage, preferences are estimated using choice data and the first stage estimates. However, our reasons for using a two stage approach are

[2] Varian (1988) proves a similar result but with linear prices and finitely many observations per consumer.

slightly different from those of Rosen (1974). Rosen's (1974) primary use of the first stage is to assist in estimating second stage utility. We also use the first stage estimates in the second stage, but we primarily need the first stage to identify and estimate the unobserved product characteristics.

The first stage of our estimation procedure relies on the existence of a hedonic pricing function. In section 3 we prove that if demand is given by the hedonic discrete choice model, then prices in each market can be written as a function of the observable and unobservable characteristics of the products in that market. We also show, under some weak conditions, that the equilibrium price surface must be Lipschitz continuous in characteristics, and strictly increasing in the unobserved characteristic.

Since even the simplest I.O. models of competition would suggest that any equilibrium price function should be nonlinear and nonseparable in all the product characteristics, it would not be appropriate to specify the unobservable product characteristics as additively separable and then proceed using standard econometric techniques such as OLS or IV. Instead, we allow for a completely general functional form for the equilibrium pricing function. We show that it is possible to use price data to nonparametrically identify the unobserved product characteristics and the price function in at least four cases.

The first case we study is when unobserved product characteristics are independent of the observed product characteristics (using the results of Matzkin (1999)). This is the case that the previous literature on differentiated products demand estimation has concentrated on (e.g. BLP). Because we believe the independence assumption to be strong, we also consider several alternatives. In the second case, 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. A third case that guarantees identification requires monotonicity and rich data on prices, but is unique in requiring nor estrictions on the joint distribution of observed and unobserved product characteristics.

A fourth case is a nonseparable instrumental variables approach due to Imbens and Newey (2001). These four cases are analogous to standard approaches in separable models, corresponding to i) OLS, ii) model fixed effects, iii) product fixed effects, and iv) IV.

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 approach to be quite general, extending to both static and dynamic equilibrium contexts. The benefit of this generality is that it allows consistent estimation of preferences even in complex supply environments, and not subject to any specific assumptions about supply. The cost is that the estimation strategies we describe for the first stage, which are nonparametric, are data intensive. However, we believe that it would be straightforward to add supply side assumptions to our model, such as those in Bresnahan (1987) or Berry (1994), which would provide greater efficiency in estimation, at the cost of losing consistency in the event that the supply assumptions are false.

We also generalize the model to allow for an additive measurement error in prices, because we believe that often 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 unobserved product characteristics are identified. Section 3 of the paper is also a contribution to the literature on estimating hedonic pricing functions when there are unobservable product characteristics.

Once the unobservable product characteristics are known, we consider the identification of preferences for several alternative types of data sets on consumer choices. The problem of identification of preferences is well understood in the event that the data contains enough observations per consumer that consumers' entire demand functions are known (see Mas-Collel (1977)). 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 aggregate data.[3] In such cases, if the choice set is continuous, and if the economist can consistently estimate

[3] 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.

the hedonic price surface that relates product characteristics to prices, then for each choice observed for a given consumer it is possible to recover the consumer's marginal rate of substitution at the chosen bundle. Knowledge of the marginal rate of substitution at a single bundle is not enough to identify preferences in complete generality. However, we derive conditions under which knowing the parametric form of the utility function provides enough information to allow the econometrician to infer each consumer's entire utility function using only this information.

In section 4 we show that for many commonly used parametric forms for utility there exists an inversion between a consumer's taste parameters and the consumer's choice. This inversion allows the economist to recover each consumer's preference parameters, and then nonparametrically estimate the distribution of random coefficients (conditional on the functional form of utility). If micro data linking demographics to choices is available, then the joint distribution of random coefficients and demographics is also identified. Section 4 also shows that if multiple observations per consumer are available, or if the researcher is willing to impose a small amount of homogeneity across individuals, then higher order approximations of consumers' utility functions are possible. The latter approach is similar in spirit to that of Blundell, Browning, and Crawford (2001).

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 are not typically identified even if the parametric form of utility is known. Instead, each individual's taste coefficients can be shown to lie in a set. We also show that, under certain conditions, this set tends to be smaller when there are more products in the market, and 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.

While our two-stage approach is primarily motivated by identification concerns, an advantage of the two-stage approach is that it facilitates estimation. The two-stage estimators, which are described in detail in section 5, are computationally simple. The two-stage approach allows the use of two different data sets for the two stages, which we believe will often be desirable.

1.1 Previous Literature

This paper builds on several literatures in microeconomics and applied microeconomics. Probably the most similar literature to our paper 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), as well as a large literature of empirical work. The primary differences between this literature and our approach are, first, that we allow for some product characteristics to be unobserved, and second, that we retain complete heterogeneity in our second stage, using an inversion to generate preferences rather than a regression. Our approach to the second stage also solves the simultaneity problems that are the focus of Epple (1987) and Bartik (1987). A more detailed comparison with this literature is provided in section 6.

This paper also builds on the recent literature in I.O. on estimating discrete choice demand models with random coefficients and unobserved product characteristics, including Ackerberg and Rysman (2000), Berry (1994), Berry and Pakes (2000), BLP, Berry, Levinsohn, and Pakes (1998), Davis (2000), Goettler and Shachar (1999), Hendel (1999), Nevo (2000), Petrin (1998), McCulloch and Rossi (1996), and others. This literature, which has attempted to find better ways to estimate demand systems in markets with differentiated products, has generalized standard discrete choice demand 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. BLP in particular have shown that random coefficient logit models imply more reasonable substitution patterns than standard models. Second, this literature has emphasized the importance of product characteristics that are not observed by the econometrician (but that are observed by consumers). Several authors (e.g. Berry (1994), BLP) have shown that if unobserved product characteristics are positively correlated with price then 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.

We maintain the two generalizations listed above by studying a model in which preferences are heterogeneous and not all product characteristics are observed by the economist. Our differences with this literature are three-fold. First, we use an alternative model of the error terms (cf. Berry and Pakes (2000)). Second, we show what role parametric assumptions play in identification of the model and when they can be eliminated. Lastly, random coefficient discrete choice models with unobserved product characteristics are computationally burdensome to estimate, particularly when the number of choices becomes large.[4] Our approach is computationally light, but data intensive.

Our results also rely heavily on several other literatures. Our estimation approach in section 5.3 builds on the Bayesian analysis of discrete choice models, especially Albert and Chib (1993), Geweke, Keane, Runkle (1994), and McCulloch and Rossi (1996). We also draw on the literature on revealed preference and integrability, for instance Richter (1966), Hurwicz and Uzawa (1971), Mas-Collel (1977), 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), Imbens and Newey (2001), and particularly Matzkin (1999).

2 The Model

Let j represent a product and let $\mathcal{J}$ be the set of all products. If $\mathcal{J}$ is finite, we let $J = \#\mathcal{J}$. In our model, a commodity is a collection of a finite number of attributes that we represent

[4] 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. 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.

as a vector of real numbers. In most applications, the economist does not observe all of the product attributes relevant to the consumer. Therefore, we assume that the economist perfectly observes the first K attributes, which we denote by the vector $x_j = (x_{j1}, ..., x_{jK})$, but does not observe the attribute ξ_j. Our analysis is limited to the case where only one product attribute is not observed.[5] All attributes are perfectly observed by the consumer. We denote the set of product attributes as $X \subseteq R^{K+1}$.

We assume that there are T markets. We let $\mathcal{I}_t$ be the set of all consumers in market $t \in T$ and we index a single consumer by $i \in \mathcal{I}_t$. Consumers are utility maximizers who select a product $j \in \mathcal{J}$ 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:

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

Consumer i in market t solves the following maximization problem:

$$\max_{(j,c)\in B(y_i,t)} \quad u_i(x_j, \xi_j, c) \tag{1}$$

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

$$\widetilde{h}(y_i, t) = \{(j, c) \in \mathcal{J} \times R_+ : (j, c) \text{ solves } (1)\} \tag{2}$$

Definition. We say that $\widetilde{h}(y_i, t)$ is generated by $u_i(x, \xi, c)$ if $\widetilde{h}(y_i, t)$ satisfies (2).

The goal of this paper is the recovery of consumers' utility functions, u_i, using standard data sets.

[5] Our analysis can be generalized to an arbitrary vector of unobserved attributes, but only for the special case where these attributes are collapsible to a single index in consumers' utility functions.

2.1 Discussion

Implicit in the notation above is the assumption that products are readily identifiable, meaning that it is possible to identify two products in two different markets that are identical. We make this assumption because in our experience with choice data it has typically been the case. For example, even when automobile models retain the same name from one year to the next ("Ford Taurus"), it is usually easy to determine whether or not the product has actually been redesigned. Under this assumption, all product characteristics, including the unobservable product characteristic, ξ_j, are fixed across markets for a given product.

On the other hand, except where otherwise noted, we will not in general assume that individual preferences or the distribution of individual preferences are constant across markets. This is one of the main differences between this paper and the recent I.O. literature. Our main reason for not making this assumption is that we think that it is not likely to hold in many data sets due to the importance of complimentary goods. For example, consumers' preferences for computers change over time with the available software, and consumers' preferences for the characteristics of autos change over time with the price of fuel.[6] Assuming that preferences are constant over time would provide additional identifying information. We leave a full investigation of such restrictions to future research.

2.2 Standard Econometric Approaches

The standard econometric approach to estimating preferences has been to apply discrete choice models such as logit and probit. Typically, practical implementation of these models involves making a number of parametric, homogeneity, and independence assumptions. One of our reasons for writing this paper is a desire to understand the importance of these assumptions.

[6] One possible exception to this rule is high frequency data such as scanner data.

For example, the typical fixed coefficient discrete choice model can be written as:

$$u_{ij} = x_{1j}\beta + \epsilon_{ij} \tag{3}$$

where x_{1j} is a vector of observable characteristics, β is a taste parameter that is constant across individuals, and ϵ_{ij} is an error term with (usually) known distribution.[7] An alternative but equivalent way of writing the model is:

$$u_{ij} = x_{1j}\beta + x_{2j}\eta_i \tag{4}$$

where x_{2j} is a J-dimensional vector of zeros with a one in the jth element (a vector of product dummies), and η_i is a J-dimensional taste vector with known distribution. Thus, relative to the general model of (1), the assumptions made in (4) include a parametric form for u_i, inclusion of product dummies in x, preferences for x_{1j} that are constant across individuals, and a known distribution for η_i. Furthermore, the η_i vector is often assumed to be mutually independent as well as independent across individuals.

Many recent papers in the I.O. literature have used more flexible models. As an example, we consider the model of BLP,[8]

$$u_{ijt} = x_{1jt}\beta_i + \xi_{jt} + \alpha \log(y_i - p_{jt}) + \epsilon_{ijt} \tag{5}$$

where ξ_{jt} is an unobservable product characteristic and y_i is income. An alternative but equivalent way of writing this model that puts it in the form of (1) is,

$$u_{ijt} = x_{1jt}\beta_i + x_{2j}\eta_{it} + \beta_t\xi_{jt} + \alpha \log(y_i - p_{jt}) \tag{6}$$

We have also added a time-varying taste, β_t, for the unobservable product characteristics to show that the model is general to this possibility. Additional assumptions typically made in (6) include: β_i is normally distributed with unknown mean and variance; the distributions of y_i and of ϵ_{ijt} are known; β_i, y_i, and ϵ_{ijt} are independent of one another, as well as mutually independent. In addition, preferences for the observed characteristics are assumed

[7] We use "known distribution" to mean that the entire distribution function is known, including any parameters to a parametric form.

[8] Berry, Levinsohn and Pakes (1998), Davis (2000), Goettler and Shachar (1999), Hendel (1999), Nevo (2000), Petrin (1998), and others use variations of this model.

to be constant over time, preferences for the unobserved characteristic are constant across individuals, and preferences for the outside good are constant across individuals conditional on income.

One of the main goals of this paper it to understand which of the assumptions that are commonly made are necessary for identification and which are not, in addition to showing nonparametric identification of preferences where it exists.

One difference between the general model of (1) and the two examples above is that there is no error term in (1). As we have shown above, one way to view the error terms in standard models is as equivalent to including a set of product dummies as observable characteristics in (1). Thus, our model is in principle general to this case. However, we do not want to emphasize this generality too much since in the analysis that follows it will become obvious that if the product dummies are included for all products then identification of the unobservable product characteristics and of consumer preferences is very weak without making the kind of distributional and independence assumptions that this paper is trying to relax.

In support of our approach, we do not believe that it is in general appropriate to treat product dummies as product characteristics, as this leads to some properties of the economic model that may be unbelievable and undesirable (see our other working paper Bajari and Benkard (2001) for a full discussion, as well as Ackerberg and Rysman (2001), Anderson, de Palma and Thisse (1992), BLP, Berry and Pakes (2000), Caplin and Nalebuff (1991), and Petrin (1998) for similar arguments). Their widespread use through the error terms in standard models seems largely to have been for econometric convenience. In addition, the fact that identification of preferences is very weak when product dummies are included means that in practice the results are likely to be driven by the strict independence and distributional assumptions that are almost sure to be false.

2.2.1 Data

We consider identification and estimation of preferences in four cases:

1. **Aggregate Data:** Prices, p_{jt}, quantities, q_{jt}, and characteristics, x_j, are observed for $j = 1..J$ products and $t = 1..T$ markets.

2. **Micro Data:** In addition, for each purchase occasion, a vector of demographics of the consumer, z_i, is observed. However, each consumer is only observed once.

3. **Scanner Data:** In addition, multiple purchase occasions are observed for each consumer.

4. **Limit Case:** In addition, each consumer's choices are observed in a large number of markets and under sufficient price variation such that the consumer's entire demand function is known.

The main difference between the four cases is that they contain increasing amounts of information about consumer preferences from 1-4. The first three represent commonly available data sets. The fourth is a limiting case of the scanner data case.

We note that the observed quantities may be unity, as would be the case with, e.g., housing data. We also assume, for convenience, that all products are observed in all markets. However, that assumption is not necessary for any of the arguments in this paper.

We will also consider two types of markets: a continuous choice set, and a discrete choice set. The continuous choice set case refers to the case where the consumer can essentially choose any combination of characteristics she desires within a compact set. We think that this case is relevant to markets such as housing or computers, where the choice set is very rich. The discrete choice set case refers to the case where there are only a small number of products in the market and hence the consumer may not be able to choose a bundle that is close to her optimal bundle in characteristics space.

2.3 Non-Identification in the Best Case

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, little formal development of the theory of revealed preference for differentiated products when the economist cannot observe all of the product characteristics (exceptions include Varian (1988) and Epstein (1982)). 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 with a finite number of observations on a consumer's choices (corresponding to case 3 above) provide essentially no information about the consumer's preferences. In this section we similarly establish that it is not possible to recover consumer i's weak preference relation based on observing her choices even if the choice set is continuous and the consumer's entire demand function is known (case 4 above).

The intuition behind the result is straightforward. 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 generates the observed demands.

We begin with some standard assumptions. For a more detailed discussion, see the appendix. We assume that there are a continuum of products. For convenience, we define the set of products, $\mathcal{J}$, as the unit interval, $\mathcal{J} = [0, 1]$. We assume that the product space, X, is a convex, compact subset of $\mathbb{R}^{K+1}$ with $0 \in X$. For all t, we assume that $p_t(x, \xi)$ is a continuous function, with $p_t(0) = 0$ and $p_{ct} > 0$. The last two assumptions imply that the budget set is compact. We also assume that $u_i(x, \xi, c)$ is continuous in all its arguments, which guarantees that the demand correspondence is non-empty.

In revealed preference, it is often convenient to work with preference relations instead of utility functions since utility functions are never uniquely determined. In our model, as in the standard analysis of the consumer in partial equilibrium, maximization implies that the strong axiom of revealed preference is satisfied.

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 unobserved product characteristics 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 the consumer's choices. We define identification formally below.

Definition 1. *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 now show that, without further assumptions, the weak preference relation is not identified. This result holds so long as there is at least one point at which the utility function is strictly increasing in the observed product characteristics.

Theorem 1. *Suppose that the demand correspondence $\widetilde{h}(y_i, t)$ is generated by the utility function u_i. Also suppose that there exists at least one point $(\overline{x}, \overline{c})$ such that $u_i(x, c)$ is strictly increasing in some neighborhood of $(\overline{x}, \overline{c})$. Then the weak preference relation $\succeq_i$ is not identified.*

Sketch of Proof: Set $\xi_j = j$. Define the price function to satisfy $p_t(x_j, \xi_j) = p_{jt}$ for all x_j.[9] Since we have constructed price functions p_t that match the data and are nowhere strictly increasing in the observed product characteristics, the hedonic pricing function and unobserved characteristics are not identified.[10]

Using the construction above, suppose that all of the price is due to the product unobservable.

[9] Please see Theorem 2 for proof that a price function exists in our model.
[10] This result is obvious as it is analogous to running OLS with no assumptions on the errors.

Since the demand correspondence is generated by a utility function, it satisfies the strong axiom of revealed preference. Since demand obeys the strong axiom of revealed preference and the budget set depends only on the unobservable characteristic, it is possible following standard arguments provided in Richter(1966) to construct a preference relation over only 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, which proves that the weak preference relation is not identified. □

The above theorem demonstrates 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. 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 the 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.

2.4 Identification/Estimation Strategy

Our identification and estimation strategy in this paper is based upon a two-stage approach in the spirit of the approach suggested in Rosen (1974). In the first stage, the price function and product unobservables are estimated using data on prices and characteristics,

$$p_{jt} = p_t(x_j, \xi_j). \tag{7}$$

In the second stage, preferences u_i are estimated using choice data.

We will show in section 4 that even if all product characteristics are known (the unobserved product characteristics having been estimated in the first stage), and the parametric form of the utility function is known, the second stage of the model is just identified. Therefore, the choice data contains no additional information that can be used to estimate the unobserved product characteristics.

In order to identify the unobserved characteristics, we use additional information contained in prices in the first stage, specifically, the information that there exists a price function.[11] The next section will prove the existence of a price function under weak conditions. There is no benefit to joint estimation of the two stages because of the lack of overidentifying information in the choice data.

3 Identification of the Price Function and the Unobservable Characteristics

3.1 The Price Function

We first 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.

We make the following three assumptions.

A1 $u_i(x_j, \xi_j, c)$ 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 $c \in (0, y_i]$.

A2 u_i is Lipschitz continuous in (x_j, ξ_j).

A3 u_i is strictly increasing in ξ_j.

Assumption A3 is the most restrictive assumption of the three. It implies that there is no satiation in the unobservable product characteristic. Without A3 it would not be possible to identify the unobservable product characteristics as no inversion would exist.[12]

[11] Additional assumptions are also necessary. See section 3.

[12] For the same reason, assumption A3 is also necessary for the inversion of Berry (1994) and BLP.

Theorem 2. *Suppose that A1-A3 hold for every individual in every market. 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. □

The intuition for the theorem is that if properties (i)-(iii) were not satisfied by the equilibrium prices, then some of the goods could not have positive demand.

We denote the equilibrium price function for market t as $p_t(x_j, \xi_j)$. It is a map from the set of product characteristics to prices that satisfies $p_t(x_j, \xi_j) = p_{jt}$ for all $j \in \mathcal{J}$ and we assume throughout the rest of the paper that (i)-(iii) hold. Because (iii) holds for all pairs of products, in the limit the price function must be Lipschitz continuous.[13]

We note here that 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. This is the primary reason for the fact that the price function is different in every market (hence the subscript t). What the price function in a particular market does tell us is the relationship between characteristics and prices as perceived by a consumer in that market.

[13] Differentiability of the price function would also be desirable because it would improve the efficiency of estimation. However, it is not possible to show differentiability using purely demand based arguments. The reason for this result is that a kinked budget set set does not necessarily rule out positive demand everywhere. Differentiability of the price surface would instead have to be derived from both demand and supply side primitives. For example, if the cost function were continuously differentiable and the market were perfectly competitive then the price function would be continuously differentiable.

Because the theorem is based on demand side arguments only, these results are general to many types of equilibria, both dynamic and static. We thus 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 are 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 emphasize that even the simplest I.O. models of competition would suggest that any equilibrium price function should be nonlinear and nonseparable in all the characteristics. For example, the standard single product firm inverse elasticity markup formulas imply a nonseparable price function even if the marginal cost function is linear in characteristics. Thus, it would not be appropriate for us to specify the unobservable product characteristics as additively separable and then proceed using standard econometric approaches such as OLS or IV. Instead, we proceed by allowing the price function in each market take on a completely general form.

3.2 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,

$$p_{jt} = p_t(x_j, \xi_j), \tag{8}$$

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,

A4 ξ is independent of x.

A5 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 A4 and A5 are the primary identifying assumptions. A5 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. A5 is implied (without the lower bound) by A3 (see Theorem 2).

Let the set I be the set of price functions satisfying A5, and Γ be the set of distribution functions that are strictly increasing.

$$I = \{p' : A \times E \to \mathbb{R} \mid \text{for all } x \in X,\ p'(x, \cdot) \text{ is strictly increasing}\} \tag{9}$$

$$\Gamma = \{F : \mathbb{R} \to \mathbb{R} \mid F \text{ is strictly increasing}\} \tag{10}$$

Since the unobserved product attribute has no inherent units, it is only 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,

$$M = \{p' : A \times E \to \mathbb{R} \mid p' \in I \text{ and } p'(\bar{x}, \xi) = \xi\} \tag{11}$$

In the theorem below, we assume without loss of generality that $p_1(\cdot) \in M$.[14] Define identification to mean identification within the set satisfying the normalization made above,

Definition 2. *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 3. *If prices are observed without error, A4-A5 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.*

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,

$$p_t(x_0, e_0) = F^{-1}_{p_t|x=x_0}(F_\xi(e_0)), \tag{12}$$

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,

$$\xi_j = F^{-1}_\xi(F_{p_1|x=x_j}(p_{j1})) \tag{13}$$

□

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.

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

We now consider the case where prices are observed with error. Specifically, we assume that p_{jt} is not observed. Instead, the econometrician observes y_{jt}, where

$$y_{jt} = p_{jt} + \epsilon_{jt} \equiv p_t(x_j, \xi_j) + \epsilon_{jt} \tag{14}$$

We also assume classical measurement error.

A6 ϵ_{jt} is *iid*, and $E[\epsilon|x, \xi] = 0$.

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.

Theorem 4. *If prices are observed with error, A4-A6 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

$$\bar{p}^T(x, \xi) = \frac{1}{T} \sum_{t=1}^{T} p_t(x, \xi) \tag{15}$$

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.[15]

For each product, j, define the set

$$\mathcal{J}_j = \{k \in \mathcal{J} |\ x_k = x_j \text{ and } \lim_{T \to \infty} \bar{p}_j^T - \bar{p}_k^T = 0\} \tag{16}$$

The set $\mathcal{J}_j$ 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 $\mathcal{J}_j$:

$$p_{jt} = E[y_{kt} | k \in \mathcal{J}_j] \tag{17}$$

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

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 3 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 are also identified. $\square$

3.3 Identification Using "Options Packages"

We believe the independence assumption made in the previous subsection to be strong. Therefore, in this section we provide an alternative set of assumptions that also provide identification and that we believe may be satisfied in many applications.

In many applications, consumers may simultaneously choose 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 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 unobservable product characteristics.

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 is a partition of $\mathcal{J}$. We let z denote a model and $\mathcal{Z}$ denote the set of all models. We assume that there exists a map $\pi : \mathcal{J} \to \mathcal{Z}$ that associates products with models. The inverse image of z under the map π are those products that are model z, although with possibly different options packages. We assume that z is observable, and that x and z have joint distribution $F_{x,z} : A \times \mathcal{Z} \to \mathbb{R}$.

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

A7. For all $j_1, j_2 \in \mathcal{Z}$, 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.[16] We formalize this requirement using the following assumption,

A8. There exists an $\bar{x} \in A$ such that for all $z \in \mathcal{Z}$, $f(\bar{x}|z) > 0$.

The baseline package here corresponds to the "reference" package in the above section. Due to the lack of implicit units for ξ, we again can only 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

$$\Gamma' = \{F : \mathbb{R}^{K+1} \to \mathbb{R} \mid F \text{ is strictly increasing in the natural ordering of } \mathbb{R}^{K+1}\} \quad (18)$$

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

Theorem 5. *If prices are observed without error, A5 and A7-A8 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,

$$\begin{aligned} \xi_j &= p_1(\bar{x}, \xi_j) && (19) \\ &= p_{1k} \text{ for } k \in \pi^{-1}(\pi(j)) \text{ such that } x_k = \bar{x}. && (20) \end{aligned}$$

[16] This support condition is the price paid for non-separability.

Equation (19) holds due to the normalization. A baseline product, k, exists for every model $\pi(j)$ by A8. 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 $(x_0, e_0) \in A \times E$,

$$p_t(x_0, e_0) = p_{kt} \text{ for } k \in \mathcal{J} \text{ such that } \xi_k = e_0 \text{ and } x_k = x_0 \tag{21}$$

□

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

Theorem 6. *If prices are observed with error, A5-A6 and A7-A8 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 $\mathcal{J}_j = \{k \in \pi^{-1}(\pi(j)) \mid x_k = x_j\}$. As above, $\mathcal{J}_j$ is the set of all products with the same characteristics as j. Then

$$p_t(x_j, \xi_j) = E[y_{kt} | k \in \mathcal{J}_j], \tag{22}$$

where the measurement error again averages to zero. The rest of the proof is by Theorem 5. □

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.

3.4 Identification With a Rich Set of Price Functions

The third approach to identification is unique in that it requires no additional assumptions on the joint distribution of x and ξ. Instead, we rely on two assumptions about the set of price functions that are observed in the data. First, we suppose that the data is rich enough that there is one market in which prices do not depend very much on the observed characteristics,

A9 There exists a market, t, such that $p_t(x, \xi) = f(\xi)$, with $f_\xi > 0$.

We do not assume that the researcher knows which market this is. We also need a monotonicity condition on the price functions that are observed,

A10 For all markets t, $p_t(x, \xi)$ is weakly increasing in all of the observed characteristics, x, and strictly increasing in the unobserved characteristic, ξ.

We think that A10 is likely to hold in many applications. If all individuals have monotone preferences over all characteristics, then A10 holds by an argument similar to that of Theorem 2. However, A10 might hold even if this were not the case. For example, if marginal costs were sufficiently increasing in all characteristics, then A10 would also hold.

Theorem 7. *If prices are observed without error, A9 and A10 hold, $p_1 \in M$, and (x, ξ) have full support on $A \times E$, 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. Let $x \equiv (x_1, ..x_k, \xi_x)$ and $y \equiv (y_1, ..., y_k, \xi_y)$ be two points in the commodity space. In order to prove that the $\{\xi_j\}$ are identified, we will first show that the ranking of ξ_x and ξ_y is uniquely determined. Let $x^* = (\min(x_1, y_1), ..., \min(x_k, y_k))$ be the component by component minimum of the observed characteristics of the two products. Define $\mathcal{J}' \subseteq \mathcal{J}$ as follows:

$$\mathcal{J}' = \{j' \in \mathcal{J} \ : \ (x_{j',1}, ..., x_{j',k}) = x^* \text{ and } p_{j',t} \leq p_t(x) \text{ for all } t\} \tag{23}$$

It follows from A9 and A10 that there exists an element $j' \in \mathcal{J}'$ and a market t such that $p_{j',t} > p_t(y)$ if and only if $\xi_x > \xi_y$.

This identifies the ranking of $\{\xi_j\}$. The normalization M thus identifies the $\{\xi_j\}$ and $F_{x,\xi}$. Identification of $p(x, \xi)$ follows directly. $\square$

3.5 Identification Using Instruments

A fourth approach is provided by Imbens and Newey (2001), which demonstrates that it is possible to use an instrumental variables approach even in nonseparable models. We omit the details here in the interest of brevity. Therefore, if one can find a set of instruments that influence the observed product characteristics but that are independent of the unobserved product characteristic, then it is possible to nonparametrically identify the price surface and the unobservable product characteristics. However, we note that such instruments may be hard to obtain in practice.

4 Identification of Preferences

4.1 Identification of Preferences in the Best Case

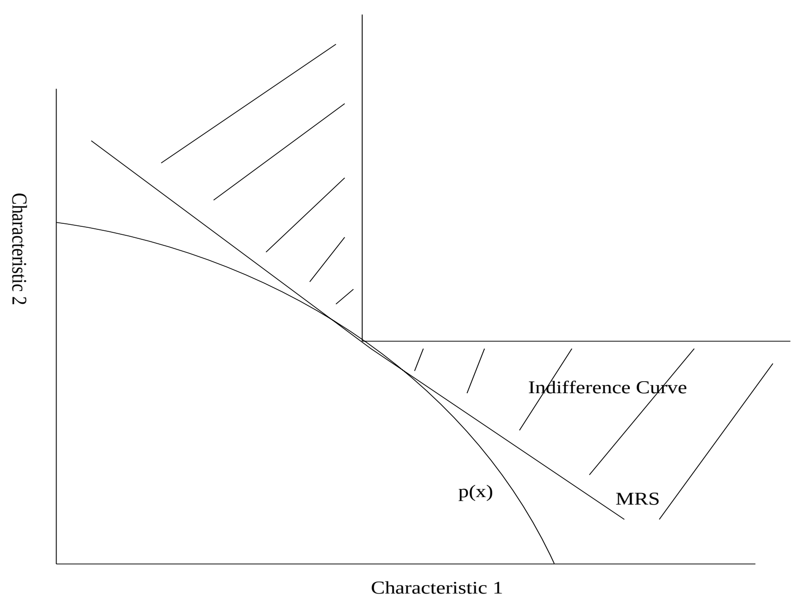
The results of section 3 demonstrate that the price functions $p_t(x, \xi)$ and the unobserved product characteristics $\{\xi_j\}$ are identified (up to a normalization). Once ξ is known, recovering the weak preference relation $\succeq_i$ from observed choice behavior is a well studied problem if there is sufficiently rich data that the consumer's entire demand function is known (case 4). We refer readers to Mas-Colell (1977) for a set of conditions under which the consumer's weak preference relation is identified.

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

In the previous sections, we studied the identification problem in the case where the economist can observe the consumer's 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 can identify the consumer's preferences. However, in applied work the economist typically observes only a handful of choices per consumer, often just one. In this section, we study identification in cases 1-3, which better reflect available data sets.

When the entire demand function is not observed, it is clear that recovery of the entire weak preference relation is not possible. In figure 4.2 we suppose that the good is a bundle of two characteristics. The slope of the budget surface identifies the marginal utilities at the chosen bundle. However, without further assumptions, it is clear that many indifference curves would rationalize the observed choice.

Figure 1: Global Identification of Indifference Curves



Some weak assumptions 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. One approach to measuring the effects of a policy change would be to use these two functional forms as bounds. However, depending on the policy of interest, we may still be left with a wide range of possibilities.[17]

The simplest way to narrow down the range of possibilities is by using functional form assumptions. Many discrete choice models used in the previous literature specify the utility function as being linear in product characteristics and the composite commodity. 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 nonparametric identification of the population distribution of random coefficients (conditional on the functional form of the utility function) even if individuals are only observed only 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:

$$u_i(x, c) = u(x, y_i - p(x); \beta_i). \tag{24}$$

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

[17] Note that if the budget set is not convex, tighter bounds can be obtained because the budget set itself is a lower bound to the indifference curve.

the following first order conditions are necessary for maximization:

$$\frac{\partial}{\partial x_k} \{u(x, y_i - p(x); \beta_i)\} = 0 \text{ for } k = 1, ..., K \tag{25}$$

The next theorem derives formal conditions under which there exists an inversion between an individual's choices and the individual's preference parameters. For simplicity, we consider the case where individuals are observed once. The case where individuals are observed more than once is a straightforward extension.

We assume that the econometrician observes the choices 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 (25) for agent i is:

$$D_\beta \left(D_{x'} \{u(x, y_i - p(x); \beta_i)\} \right) \tag{26}$$

Theorem 8. *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 (26) 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. □

Theorem 8 places tight restrictions on the types of utility functions that can be identified using the choice data. Conditional on knowing the price surface $p(x)$, we can identify at most K random coefficients per choice observation. The more observations we have for an individual (e.g., case 3: scanner data), the more flexibly we can estimate her utility function.

While this may seem like a negative result, even a first order approximation to the utility function should provide accurate results for local changes in utility. 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.

We also emphasize that, while the recent literature on discrete choice has focused on the case of linear indifference curves, this assumption is not necessary, and in many cases would be undesirable. Linear indifference curves imply an extreme amount of substitutability. In many cases, particularly those relying on the global properties of indifference curves, it may be more reasonable to use functional forms such as Cobb-Douglas, which impose declining marginal rates of substitution. Robustness checks on the results of interest subject to changes in the functional form of $u(\cdot)$ would also be desirable.

While the theorem does place restrictions on what functional forms can be identified, conditional on the functional form of utility, the entire joint distribution of the taste coefficients is identified even in the case of aggregate data (case 1). If demographics are observed (case 2), then the joint distribution of demographics and taste coefficients, $F(z, \beta)$, is identified. In the case of aggregate data, demographics are not observed so the joint distribution of demographics and tastes is not identified. However, if the distribution of demographics and tastes are assumed to be independent, and the marginal distribution of demographics is observed, then the joint distribution is identified even in aggregate data. This is the strategy used in the recent I.O. literature.

Finally, we note that if the number of random coefficients to be estimated equals the number of characteristics ($B = K$), as would typically be the case, then the distribution of the random coefficients is just identified. All of the information in the choice data is needed in order to estimate preferences, and there is no information remaining.

4.3 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 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 (25) 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

$$u_{ij}(\beta_i, x_j, y_i - p_j) \geq u_{ik}(\beta_i, x_k, y_i - p_k) \text{ for all } k \neq j \tag{27}$$

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

$$A_{ij} = \{\beta_i : u_{ij}(\beta_i, x_j, y_i - p_j) \geq u_{ik}(\beta_i, x_k, y_i - p_k) \ \forall k \neq j\}. \tag{28}$$

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 small. We show in appendix section 8.4 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 5.3, we follow the latter strategy. In either case, the identification of the aggregate distribution is weakest in the cases of (2) and (3) above.

Note that, while there is no need to explicitly model the price function in the discrete choice set case, the price function plays an identical role here to the continuous choice set case. The solution to the discrete maximization problem will be close to the solution of the continuous maximization problem if the choice set is rich. In that case, even if the derivatives or the price function are never estimated, the A_{ij} set will be close to the set of parameters that solve the continuous first order condition. Implicitly, the solution is the same even though it is obtained in a different way. Therefore, even in the discrete choice set case it is important to take care in choosing the global functional form for the utility function, subject to the discussion in section 4.2.

4.4 Imposing Homogeneity

Up to this point, we have not used any information in the data across individuals. In the aggregate and micro data cases, where there is only a single observation per individual, without homogeneity restrictions, certain features of preferences such as the Engel curve for each individual are not identified except through functional form assumptions.

Suppose we were to find in the data that the taste coefficients are correlated with income. This correlation could be interpreted in two ways. One possibility is that people born into rich families have different tastes than those born into poor families, and that their income is correlated with their parents income. Another possibility is that tastes change in a systematic way with income, i.e., if we were to give a poor family more income their tastes would change to look more like those of rich families.

Since the latter explanation has some appeal, it may in certain cases be preferable to incorporate this feature into our model. One way to do this is to impose some homogeneity across individuals.

Assume that for some group of individuals,

$$\beta_{ik} = f_k(y_i) + \eta_{ik} \qquad \text{where } E[\eta_{ik}|y_i] = 0 \tag{29}$$

Equation (29) uses covariation in income and tastes across individuals to identify individual's Engel curves. However, it retains the differentiation in tastes across individuals. Because η_i is held fixed when income changes, an individual with a strong preference for a certain characteristic relative to other individuals with similar income always has a stronger than average preference for that characteristic, regardless of her income.

Another way to interpret (29) is that it is using homogeneity restrictions to obtain a higher than first order approximation to the utility function. With this interpretation, our approach is similar in spirit to that of Blundell, Browning, and Crawford (2001).

4.5 Parametric Identification

In both the continuous choice set case and the discrete choice set case, it is straightforward to estimate the distribution of preferences parametrically using, e.g., maximum likelihood. In both cases, parametric forms are overidentified so long as there are more products in the market than there are parameters.

5 Estimation

5.1 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 maintain all of the assumptions in section 3.2. 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.

In the discrete choice set case (section 5.3 below) our first stage consists of using prices to estimate the value of the unobservables. In the continuous choice set case, it is also necessary to know the price function derivatives. If there is measurement error, then before the first stage estimation 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.

5.1.1 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,

$$y_{jt} = p_t(x_j, \xi_j) + \epsilon_{jt} \tag{30}$$

using the expectation in equation (17). To do this, as in the identification section, we 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,

$$\xi = \Psi(x, \bar{p}^T) \tag{31}$$

Let

$$g_t^T(x, \bar{p}^T) \equiv p_t(x, \Psi(x, \bar{p}^T)) \tag{32}$$

What we would like to do is to estimate equation (30). However, we cannot do that because ξ is not observed. Instead, we substitute $g_t^T(x_j, \bar{p}_j^T)$ for $p_t(x_j, \xi_j)$ and regress observed prices on the observed characteristics and average prices:

$$y_{jt} = g_t^T(x_j, \bar{p}_j^T) + \epsilon_{jt} \tag{33}$$

We 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

$$\hat{g}_t^T(x_0, \bar{p}_0^T) = \frac{\frac{1}{J}\sum_{k=1}^{J} \frac{1}{h_J^{K+1}} K_1\left(\frac{\bar{p}_0^T - \bar{y}_k^T}{h_J}\right) K_2\left(\frac{x_0 - x_k}{h_J}\right) y_{kt}}{\frac{1}{J}\sum_{k=1}^{J} \frac{1}{h_J^{K+1}} K_1\left(\frac{\bar{p}_0^T - \bar{y}_k^T}{h_J}\right) K_2\left(\frac{x_0 - x_k}{h_J}\right)}. \tag{34}$$

$\bar{y}_k^T = \frac{1}{T}\sum_{t=1}^T y_{kt}$ is an estimator for $\bar{p}_k^T$, which we need to estimate because it is not observed directly due to the measurement error. Thus, we are essentially plugging a first step parametric estimator into the second step nonparametric estimator.

We make the following assumptions (details in appendix):

B1 $K_1(\cdot)$ 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.

B2 $\lim_{J\to\infty} h_J = 0$, $\lim_{J\to\infty} Jh_J^{K+1} = \infty$, $\lim_{J\to\infty} h_J^m \sqrt{Jh_J^{K+1}} = \lambda$ where $0 \leq \lambda < \infty$.

B3 x_j, ξ_j are *iid*, mutually independent, and distributed according to $F(x,\xi)$, with density $f(x_j,\xi_j)$.

B4 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.

B5 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 $2 < r \leq \infty$.

B6 $J^{-2/r}Th_J^{\frac{2(m_2+1)}{m_2}} \to \infty$, where r is as in B5.

The non-standard assumptions are B5 and B6, which are needed in order to assure that the estimated $\bar{y}_k^T$ terms do not affect the estimation of the true prices. B6 requires that T increase 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 9. *Under B1-B6,*

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

The first step estimator affects the finite sample performance of the two-step estimator, but not the asymptotic performance so long as B5-B6 hold.

An alternative approach would be to remove the measurement error using a series estimator.

5.1.2 Estimation of $\{\xi_j\}$

Let $\hat{F}_{p1|x=x_0}(e_0)$ 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 a kernel estimator (such as those outlined in Matzkin (1999)) or a series estimator (such as those outlined in Imbens and Newey (2001)) could be used. Define an estimator for ξ by the following,

$$\hat{\xi}_j = \hat{F}^{-1}_{p_1|x=\bar{x}} \hat{F}_{p_1|x=x_j}(\hat{p}_{j1})) \qquad (35)$$

While Matzkin (1999) does not explicitly consider estimation of the unobservable, the asymptotic properties of the estimator in (35) are analogous to those of the estimator considered in Theorem 4 of that paper.

If there is measurement error, then the same estimators can be used except that it is first necessary to estimate the true prices as in section 5.1.1 above. Note that after plugging in the estimated true prices, the asymptotic properties of the estimator change. This is because the estimator in section 5.1.1 has dimension $K + 1$ while the estimator $\hat{F}$ has dimension K. Again for brevity and because much of the work would replicate results from the previous literature, we omit the asymptotic properties of the measurement error estimator here.

5.2 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 assume that the utility function takes the following form (omitting the t subscripts),

$$u_{ij} = log(x_j)\beta_{i,x} + log(\xi_j)\beta_{i,\xi} + log(y_i - p_j) \tag{36}$$

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 would be that the joint distribution of demographics and taste coefficients would not in general be identified (see section 4.2 for a discussion).

Assuming an interior maximizer, equation (36) is maximized at

$$\frac{\beta_{i,k}}{x^*_{j,k}} = \frac{1}{y_i - p(x^*_j, \xi^*_j)} \frac{\partial p(x^*_j, \xi^*_j)}{\partial x_k} \quad \text{for } k \in 1..K \tag{37}$$

$$\frac{\beta_{i,\xi}}{\xi^*_j} = \frac{1}{y_i - p(x^*_j, \xi^*_j)} \frac{\partial p(x^*_j, \xi^*_j)}{\partial \xi} \tag{38}$$

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

$$\hat{\beta}_{i,k} = \frac{x^i_{j,k}}{y_i - \hat{p}^i_j} \frac{\partial \widehat{p(x^i_j, \xi^i_j)}}{\partial x_k} \quad \text{for } k \in 1..K \tag{39}$$

$$\hat{\beta}_{i,\xi} = \frac{\hat{\xi}^i_j}{y_i - \hat{p}^i_j} \frac{\partial \widehat{p(x^i_j, \xi^i_j)}}{\partial \xi} \tag{40}$$

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, it is thus possible to estimate the vector of taste coefficients for each individual. One way to estimate the price function derivatives is by using the derivatives of a price function estimator. The price function can be estimated analogously to (35) above (except using (12)) and using either a kernel or series-based approach. Matzkin (1999) also provides a direct estimator for the price function derivatives.

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 choice observation.[18]

Using the estimated taste coefficients for a sample of individuals along with their observed demographics, it is then possible to construct a density estimate of the joint distribution of taste coefficients and demographics in the population.

5.3 Estimation of Preferences, Discrete Case

In this section, we propose an approach to estimation when the set of products is finite. As we discussed in section 3.7, the taste coefficients are typically not identified in this case. Our approach, therefore, is to recover sets of taste coefficients that are consistent with a consumer's choices as opposed to point estimates of taste coefficients. The approach we develop is in the spirit of the bounds approach (see Manski (1995, 1997) and Manski and Pepper (2000)). The numerical techniques for our analysis borrow heavily from Bayesian estimation of discrete choice models (Albert and Chib (1993), Geweke, Keane, and Runkle (1994), and McCulloch and Rossi (1996)).

We begin by considering the problem when we see a cross section of consumers and markets. That is, we see a set of spatially distinct markets, each with a distinct price for the $j = 1, ..., J$

[18] Note that a single choice observation reflects a K-dimensional choice vector.

products and we see the choice of each consumer exactly once. Let the utility of consumer i for product j , u_{ij} satisfy:

$$u_{ij} = u(x_j, y_i - p_j; \beta_i) \tag{41}$$

where x_j is a vector of characteristics, $y_i - p_j$ is consumption of all other goods and β_i is a vector of taste parameters. If consumer i chooses product j then it must be the case that product j maximizes consumer i's utility:

$$u(x_j, y_i - p_j; \beta_i) \geq u(x_k, y_i - p_k; \beta_i) \text{ for all } k \neq j \tag{42}$$

As we argued in section 3.7, in general many values of β_i satisfy the set of inequalities (42). However, these inequalities provide bounds on the set of taste coefficients consistent with choosing product j. As the number of products becomes large, the set of inequalities (42) can become quite complicated. For each consumer i, there are $J-1$ inequalities that must be satisfied. In addition, the unobserved product characteristics enter into $J-1$ inequalities for each consumer in the data set. Despite the complicated nature of these inequalities, we have found a simple numerical approach to the problem that works in a large class of models. We cast the problem of estimating the taste coefficients β_i, $i = 1, ..., I$ into a Bayesian paradigm. We construct a likelihood function and a prior distribution over the parameters such that the support of the posterior distribution corresponds to the set of parameters that satisfy equation (42). We show that there exists a straightforward Gibbs sampling algorithm to simulate the posterior distribution. As the number of simulation draws becomes sufficiently large, we can learn the support of the posterior distribution and hence the set of parameters that solve the parameters (42).

The inequalities (42) generate a likelihood function in a natural fashion. The likelihood that a consumer with taste coefficients β_i chooses product j is:

$$L(j|x, y_i, \beta_i) = \begin{cases} 1 & \text{if } u(x_j, y_i - p_j; \beta_i) \geq u(x_k, y_i - p_k; \beta_i) \text{ for all } k \neq j \\ 0 & \text{otherwise} \end{cases} \tag{43}$$

That is, consumer i chooses product j so long as her taste coefficients imply that product j is maximizing. In what follows, it is technically convenient to assume that the prior distribution for β_i, $p(\beta_i)$ has a uniform distribution over the region B. Typically this region would be defined by a set of conservative upper and lower bounds for each taste coefficient. The posterior distribution for β_i, $p(\beta_i|C(i), x, p)$ conditional on the econometrician's information set then satisfies:

$$p(\beta_i|C(i), x, p) \propto \pi(\beta_i)L(j|x, \beta_i) \tag{44}$$

The posterior distribution is uniform over those β_i that are consistent with the agents choice. So long as B completely covers all of the A_{ij} sets (see (28) 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 cover estimation of that below. In general, the object of interest can be written as:

$$\int g(\beta_i)p(\beta_i|C(i), x, p) \tag{45}$$

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

$$\lim_{S\to\infty} \frac{1}{s} \sum_{s=1}^{S} g(\beta_i^{(s)}) = \int g(\beta_i)p(\beta_i|C(i), x, p) \tag{46}$$

In what follows, we 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 (43) to find the following conditional distributions:

$$p(\beta_{i,1}|x,p,C(i)=j,\beta_{i,-1}) \tag{47}$$
$$p(\beta_{i,2}|x,p,C(i)=j,\beta_{i,-2}) \tag{48}$$
$$\vdots \tag{49}$$
$$p(\beta_{i,K}|x,p,C(i)=j,\beta_{i,-K}) \tag{50}$$

If the support of the posterior distribution is not connected, Gibbs sampling will experience problems with convergence. One way to avoid these convergence problems is through an intelligent choice of a flexible functional form for $u(\beta_i, x_j, y_i - p_j)$. If $u(\beta_i, x_j, y_i - p_j)$ is modeled using an n^{th} order polynomial, it is straightforward to demonstrate that the set of β_i that satisfy the inequalities (42) are convex and therefore connected.

If the specification of utility is linear in the β_i and X_j , it is straightforward to derive the conditional densities (47). For example, consider the following model[19] :

$$u(\beta_i, x_j, y_i - p_{j,m}) = \sum_k \beta_{i,k} \log(x_{j,k}) + \log(y_i - p_{j,m}) \tag{51}$$

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

$$\sum_l \beta_{i,l} \log(x_{l,j}) + \log(y_i - p_j) \geq \sum_l \beta_{i,l} \log(x_{l,k}) + \log(y_i - p_k) \text{ for all } k \neq j \tag{52}$$

which implies that:

$$\beta_{i,1} \geq \frac{\sum_{l\neq 1} \beta_{i,l}(\log(x_{l,k}) - \log(x_{l,j})) + \log(y_i - p_j) - \log(y_i - p_k)}{\log(x_{1,j}) - \log(x_{1,k})} \text{ if } x_{1,j} > x_{1,k} \tag{53}$$
$$\beta_{i,1} \leq \frac{\sum_{l\neq 1} \beta_{i,l}(\log(x_{l,k}) - \log(x_{l,j})) + \log(y_i - p_j) - \log(y_i - p_k)}{\log(x_{1,j}) - \log(x_{1,k})} \text{ if } x_{1,j} < x_{1,k} \tag{54}$$

[19]If we only observe the consumer choose once, even with a continuum of products, we can at best hope to identify the marginal utilities. While our approach would allow for a more flexible specification of $u(\beta_i, x_j, \xi_j, y_i - p_j)$ we see no particular gain from this more flexible parameterization when the consumer is only observed to choose once.

Since the prior distribution is uniform and the likelihood is also uniform, it follows immediately that the conditional distribution (47) must satisfy the inequalities implied by (53)-(54) and must also lie in the set B that defines the support of the prior.

To summarize, the conditional distribution (47) is uniform on the interval $[\beta_{1,\min}, \beta_{1,\max}]$ where the support satisfies:

$$\beta_{1,\min} = \max\left\{ \min_{\beta_1|\beta_{-1}} B, \max\left\{ \frac{\sum_{l\neq 1} \beta_{i,l}(\log(x_{l,k}) - \log(x_{l,j})) + \log(y_i - p_j) - \log(y_i - p_k)}{\log(x_{1,j}) - \log(x_{1,k})} \text{ such that } x_{1,j} > x_{1,k} \right\}\right\} \tag{55}$$

$$\beta_{1,\max} = \min\left\{ \max_{\beta_1|\beta_{-1}} B, \min\left\{ \frac{\sum_{l\neq 1} \beta_{i,l}(\log(x_{l,k}) - \log(x_{l,j})) + \log(y_i - p_j) - \log(y_i - p_k)}{\log(x_{1,j}) - \log(x_{1,k})} \text{ such that } x_{1,j} < x_{1,k} \right\}\right\} \tag{56}$$

The conditional distribution for the remaining β's is also a uniform distribution defined by inequalities that are analogous to (55) and (56).

Next, to evaluate the integral defined by (46) 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 (46) converges as the number of simulations tend to infinity. The support of the posterior distribution is equal to the set containing all the taste parameters β_i that are consistent with the observed choices and the a priori bounds. Even though this is potentially a complicated system of inequalities, we have found that in monte carlo experiments with over 100 products, the posterior distribution appears to converge in just a few minutes.

The algorithm defined above is 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. Suppose for instance the econometrician had access to panel data that allowed her to see each consumer choose more than once. If the preference parameter vector β_i is held fixed for each consumer, we now have just a slightly more complicated set of inequalities. Namely, the consumer's preference parameter must be such that in each time period, the consumer is maximizing utility. Upper and lower bounds for the preference parameter must then be found.

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. This can be estimated using the empirical frequency from our posterior simulations. Let $F(\beta_1, ..., \beta_K)$ be the cumulative distribution function for the K taste coefficients. It follows from (46) that:

$$
\begin{aligned}
F(\overline{\beta}_1, ..., \overline{\beta}_K) &= \Pr(\beta_1 \leq \overline{\beta}_1, ..., \beta_K \leq \overline{\beta}_K) \\
&= \lim_{S \to \infty} \frac{1}{S} \sum_{s=1}^{S} 1\{\beta_1 \leq \overline{\beta}_1, ..., \beta_K \leq \overline{\beta}_K\}
\end{aligned}
$$

A first difficulty that might be faced in practice is finding a value of $\beta_i^{(0)}$ that satisfies the set of inequalities (42). However, consider the following linear programming problem:

$$\min \sum_{i,k} \beta_{i,k} \tag{57}$$

subject to (42), and

$$\beta_{i,k} \geq 0 \text{ for all } i,k \tag{58}$$

Standard numerical packages can be used to find a solution to problem (57) and hence a starting point for the Gibbs sampling algorithm.

A second difficulty that might be faced in practice is when there are no parameter values that satisfy the inequalities in equation (42) (so that A_{ij} is empty for some i and j). This would happen if, for example, one product strictly dominated another in all dimensions of characteristics space (suppose one product was "better" in every dimension of characteristics space than another but had a lower price). Such an occurrence 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 these we would interpret that as a rejection of the functional form.

In the above algorithm, we have proceeded as if all product characteristics were observed. This will not in general be true as we have emphasized in previous sections. One approach to this problem would to use an estimate of ξ_j obtained as in section 5.1.2 and proceed as above. [20]

It can easily be shown that if all of the characteristics are continuous, then the discrete model converges to the continuous model at a rate of $\frac{1}{J}$ (see section 8.4). That is, as the number of

[20] 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:

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

products becomes large a consumer's true preference parameter is perfectly learned, and in expectation the set that contains the consumer's preference parameter has measure $\frac{1}{J}$. This is because the average measure of the A_{ij} sets is $\frac{1}{J}$.

6 Comparison with Two-Stage Hedonics

Our two-stage approach is similar to that of the two-stage hedonics literature started by Rosen (1974), with two primary differences. First, we treat one product characteristic as unobserved to the economist and allow the price function to be nonseparable in the unobserved product characteristic. Second, our second stage is an inversion rather than a regression, and thus is not subject to many of the criticisms of the second stage regression of Rosen (1974) (see Epple (1987) and Bartik (1987)). Because the first stage comparison is straightforward, we compare only the second stage in this section.

The second stage demand equation of Rosen (1974) can be represented as follows,[21]

$$\frac{\partial p(x^i)}{\partial x_k^i} = F_k(x^i, y_i) + \epsilon_{ik} \tag{59}$$

where x^i is the bundle chosen by individual i, and y_i are consumer demographics (empirical counterparts to consumer tastes). The utility function implied by this equation is

$$u_i(x, c) = u(x, y_i, c) + \epsilon' x - p(x) \tag{60}$$

where ϵ is a K-dimensional vector of the error terms in equation (59), and F_k is the kth partial derivative of $u(\cdot)$.

The problems with running regression (59) in practice are well-known (see Epple (1987), Bartik (1987), and others). The primary problem is that maximization implies that ϵ and x are correlated, and it has proven difficult to find valid instruments. Epple (1987) suggests an instrumental variables strategy that uses information across different markets. Another problem is that in practice the procedure imposes a lot of homogeneity across individuals.

[21] Rosen (1974) equation (16), page 50.

An implication of imposing this homogeneity is that the only utility function to rationalize the data (in which different individuals choose different bundles) is if all individuals have utility function $p(x)$. A third criticism, that we believe is new to this paper, is that typically the errors in (59) were not treated as structural and were thus thrown out for the purposes of economic analysis.

The two-stage procedure described in this paper solves these three problems. The second stage inversion retains a large amount of preference heterogeneity and thus naturally leads to the kind of sorting described above, yet still retains the ability to identify preferences using data for just a single market. Furthermore, the inversion is just identified so no information in the choice data is thrown away.

7 Conclusions

This paper has investigated the identification and estimation of hedonic discrete choice models of differentiated products. We showed that in general the hedonic discrete choice model with an unobserved product characteristic is not identified even if the entire demand function is observed. Moreover, we showed that even if all characteristics are observed, preferences are only just identified. We concluded that choice data contains no information about unobserved product characteristics.

However, if the unobserved product characteristic corresponds to a model, or if the unobserved product characteristic is independent of the observed characteristic, or if the data contains a rich set of price functions, or if it is possible to find instruments, we showed that it is possible to use information in prices to recover the unobserved product characteristics. These assumptions are analogous to standard econometric assumptions in separable models and we think that they are likely to hold in many applications.

Once the unobserved characteristics are known, identification of preferences is possible through revealed preference. In the random coefficient models that have been commonly used, where the parametric form of the utility function is known, there exists an inversion between a consumer's choice and her preference parameters. In such cases, knowledge of the consumers' marginal utilities at a single bundle is sufficient to non-parametrically identify the population distribution of random coefficients.

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 converges to an individual consumer's random coefficients and the population distribution of random coefficients can be recovered. The Gibbs procedure is also computationally simple.

8 Appendix

8.1 Details of Non-Identification section

Theorem 10. *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. □

We define a weak preference relation for consumer i.

Definition 3. *We say that* $\succeq_i$ *is a* ***weak preference relation*** *for consumer* i *if for all* $j, j' \in \mathcal{J}$*,* $(x_j, c) \succeq_i (x_{j'}, c')$ *if and only if* $u_i(x_j, \xi_j, c) \geq u_i(x_{j'}, \xi_{j'}, c')$*.*

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

Definition 5. *We say that* S_i *satisfies the* ***strong axiom of revealed preference*** *if* S_i *is acyclic, that is, there does not exist* $(j_1, c_1), (j_2, c_2), ..., (j_n, c_n)$ *such that:*

$$(j_1, c_1)S_i(j_2, c_2) \text{ and } (j_2, c_2)S_i(j_3, c_3) \text{ and...and } (j_{n-1}, c_{n-1})S_i(j_n, c_n) \text{ and } (j_n, c_n)S_i(j_1, c_1)$$

Theorem 11. *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. □

8.2 Proof of Theorem 2

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

$$|u_i(x_j, \xi_j, y_i - p_{jt}) - u_i(x_{j'}, \xi_{j'}, y_i - p_{jt})| \leq M_1(|x_j - x_{j'}| + |\xi_j - \xi_{j'}|) \tag{61}$$

By a mean value expansion, for all individuals,

$$u_i(x_{j'}, \xi_{j'}, y_i - p_{j't}) = u_i(x_{j'}, \xi_{j'}, y_i - p_{jt}) + (p_{jt} - p_{j't}) \frac{\partial u_i(x_{j'}, \xi_{j'}, y_i - p^*_{jt})}{\partial c} \tag{62}$$

where $p^*_{jt} \in [p_{jt}, p_{j't}]$ and varies for each i. Plugging (62) into (61) gives

$$\begin{aligned} \Big|(u_i(x_j, \xi_j, y_i - p_{jt}) - u_i(x_{j'}, \xi_{j'}, y_i - p_{j't})) + (p_{jt} - p_{j't}) \frac{\partial u_i(x_{j'}, \xi_{j'}, y_i - p^*_{jt})}{\partial c}\Big| \\ \leq M_1(|x_j - x_{j'}| + |\xi_j - \xi_{j'}|) \end{aligned} \tag{63}$$

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,

$$
\begin{aligned}
(p_{jt} - p_{j't}) \frac{\partial u_i(x_{j'}, \xi_{j'}, y_i - p^*_{jt})}{\partial c} &\leq M_1(|x_j - x_{j'}| + |\xi_j - \xi_{j'}|) \text{ for } i \text{ that prefer } j \text{ to } j'. && (64)\\
(p_{jt} - p_{j't}) &\leq \frac{M}{\frac{\partial u_i(x_{j'}, \xi_{j'}, y_i - p^*_{jt})}{\partial c}} (|x_j - x_{j'}| + |\xi_j - \xi_{j'}|) \text{ for } i \text{ that prefer } j \text{ to } j' && (65)\\
&\leq \frac{M_1}{\epsilon} (|x_j - x_{j'}| + |\xi_j - \xi_{j'}|) && (66)\\
&= M_2(|x_j - x_{j'}| + |\xi_j - \xi_{j'}|) && (67)
\end{aligned}
$$

□

8.3 Proofs for section 5.1.1

Some definitions:

Definition 6. *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 7. *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 $0 \leq I \leq p$.

8.3.1 Lemma for $\bar{\epsilon}$

Let $\bar{\epsilon}_k = \frac{1}{T}\sum_{t=1}^{T}\epsilon_{kt}$. We 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 $J^{-2/r}Th_J^2 \to \infty$, where r is as in E1.

Lemma 12. *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.

$$\begin{aligned} Pr(sup_k|\frac{\bar{\epsilon}_k}{h_J}| < \delta) &= Pr(sup_k|\bar{\epsilon}_k| < \delta h_J) & (68)\\ &= \Pi_{k=1}^{J} Pr(|\bar{\epsilon}_k| \leq \delta h_J) & (69)\\ &\geq Pr(|\bar{\epsilon}_1| \leq \delta h_J)^J & (70)\\ &\geq \left(1 - \frac{E|\bar{\epsilon}_1|^r}{\delta^r h_J^r}\right)^J & (71)\\ &= ((1 - z_J)^{1/z_J})^{z_J J} & (72) \end{aligned}$$

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 (72) 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. $\square$

8.3.2 Proofs for $\hat{g}$

Let,

$$\hat{g}_t^T(x_0, \bar{p}_0^T) = \frac{\frac{1}{J}\sum_{k=1}^J \frac{1}{h_J^{K+1}} K_1\left(\frac{\bar{p}_0^T - \bar{y}_k^T}{h_J}\right) K_2\left(\frac{x_0 - x_k}{h_J}\right) y_{kt}}{\frac{1}{J}\sum_{k=1}^J \frac{1}{h_J^{K+1}} K_1\left(\frac{\bar{p}_0^T - \bar{y}_k^T}{h_J}\right) K_2\left(\frac{x_0 - x_k}{h_J}\right)} \tag{73}$$

and let,

$$\hat{\hat{g}}_t^T(x_0, \bar{p}_0^T) = \frac{\frac{1}{J}\sum_{k=1}^J \frac{1}{h_J^{K+1}} K_1\left(\frac{\bar{p}_0^T - \bar{p}_k^T}{h_J}\right) K_2\left(\frac{x_0 - x_k}{h_J}\right) y_{kt}}{\frac{1}{J}\sum_{j=1}^J \frac{1}{h_J^{K+1}} K_1\left(\frac{\bar{p}_0^T - \bar{p}_k^T}{h_J}\right) K_2\left(\frac{x_0 - x_k}{h_J}\right)} \tag{74}$$

Then

$$\hat{g}_t^T(x_j, \bar{y}_j^T) - p_t(x_j, \xi_j) = (\hat{\hat{g}}_t^T(x_j, \bar{p}_j^T) - p_t(x_j, \xi_j)) + (\hat{g}_t^T(x_j, \bar{y}_j^T) - \hat{\hat{g}}_t^T(x_j, \bar{p}_j^T)) \tag{75}$$

Uniform Consistency:

The first term of equation (75) 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,

$$\begin{aligned}&(\hat{g}_t^T(x_j, \bar{y}_j^T)\hat{h}(x_j, \bar{y}_j^T) - \hat{\hat{g}}_t^T(x_j, \bar{p}_j^T)\hat{\hat{h}}(x_j, \bar{p}_j^T)) \\ &= \frac{1}{Jh_J^{K+1}}\sum_{k=1}^J \left[K_1\left(\frac{\bar{y}_j^T - \bar{y}_k^T}{h_J}\right) - K_1\left(\frac{\bar{p}_j^T - \bar{p}_k^T}{h_J}\right)\right] K_2\left(\frac{x_j - x_k}{h_J}\right) y_{kt}\end{aligned} \tag{76}$$

$$= \frac{1}{Jh_J^{K+1}} \sum_{k=1}^{J} \left(\frac{\bar{\epsilon}_j^T - \bar{\epsilon}_k^T}{h_J} \right) K_1' \left(\frac{\bar{p}_j^T - \bar{p}_k^T}{h_J} \right) K_2 \left(\frac{x_j - x_k}{h_J} \right) y_{kt}$$

$$+ \frac{1}{2} \frac{1}{Jh_J^{K+1}} \sum_{k=1}^{J} \left(\frac{\bar{\epsilon}_j^T - \bar{\epsilon}_k^T}{h_J} \right)^2 K_1'' \left(\lambda_{jk}^T \right) K_2 \left(\frac{x_j - x_k}{h_J} \right) y_{kt} \tag{77}$$

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 (77) gives,

$$= \frac{1}{Jh_J^{K+1}} \sum_{k=1}^{J} \frac{\bar{\epsilon}_j^T - \bar{\epsilon}_{kt}^{T-1}}{h_J} K_1' \left(\frac{\bar{p}_j^T - \bar{p}_k^T}{h_J} \right) K_2 \left(\frac{x_j - x_k}{h_J} \right) y_{kt}$$

$$- (Th_J)^{-1} \frac{1}{Jh_J^{K+1}} \sum_{k=1}^{J} K_1' \left(\frac{\bar{p}_j^T - \bar{p}_k^T}{h_J} \right) K_2 \left(\frac{x_j - x_k}{h_J} \right) y_{kt} \epsilon_{kt}$$

$$+ \frac{1}{2Jh_J^{K+1}} \sum_{k=1}^{J} \frac{(\bar{\epsilon}_j^T - \bar{\epsilon}_{kt}^{T-1})^2}{h_J^2} K_1''(\lambda_{jk}^T) K_2 \left(\frac{x_j - x_k}{h_J} \right) y_{kt} \tag{78}$$

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

$$+ (Th_J)^{-2} \frac{1}{2Jh_J^K} \sum_{k=1}^{J} K_1''(\lambda_{jk}^T) K_2 \left(\frac{x_j - x_k}{h_J} \right) y_{kt} \epsilon_{kt}^2$$

$$\leq \sup_k \left| \frac{\bar{\epsilon}_j^T - \bar{\epsilon}_{kt}^{T-1}}{h_J} \right| \frac{1}{Jh_J^{K+1}} \sum_{k=1}^{J} \left| K_1' \left(\frac{\bar{p}_j^T - \bar{p}_k^T}{h_J} \right) K_2 \left(\frac{x_j - x_k}{h_J} \right) \right| |y_{kt}|$$

$$+ (Th_J)^{-1} \left| \frac{1}{Jh_J^{K+1}} \sum_{k=1}^{J} K_1' \left(\frac{\bar{p}_j^T - \bar{p}_k^T}{h_J} \right) K_2 \left(\frac{x_j - x_k}{h_J} \right) y_{kt} \epsilon_{kt} \right|$$

$$+ \sup_k \frac{(\bar{\epsilon}_j^T - \bar{\epsilon}_{kt}^{T-1})^2}{2h_J^3} sup_\lambda |K_1''(\lambda)| \frac{1}{Jh_J^K} \sum_{k=1}^{J} \left| K_2 \left(\frac{x_j - x_k}{h_J} \right) \right| |y_{kt}| \tag{79}$$

$$+ (Th_J^2)^{-1} \sup_k \left| \frac{\bar{\epsilon}_j^T - \bar{\epsilon}_{kt}^{T-1}}{h_J} \right| \sup_\lambda |K_1''(\lambda)| \frac{1}{Jh_J^K} \sum_{k=1}^{J} \left| K_2 \left(\frac{x_j - x_k}{h_J} \right) \right| |y_{kt} \epsilon_{kt}|$$

$$+ (Th_J)^{-2} \sup_\lambda |K_1''(\lambda)| \frac{1}{2Jh_J^K} \sum_{k=1}^{J} \left| K_2 \left(\frac{x_j - x_k}{h_J} \right) \right| |y_{kt} \epsilon_{kt}^2|$$

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 (75) 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,

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

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 (75) and the bound for the second term provided by (79). By the lemma above and standard results is easy to show that under assumptions C5 and C6 the five terms in (79) 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.

8.4 Convergence of the Discrete Model to the Continuous Model

In the text, we demonstrated that typically when the set of products is discrete, the individual taste parameters cannot be uniquely recovered. However, in this section, we will demonstrate that as the number of choices become sufficiently large, under suitably regularity conditions the choice parameters can be learned in the limit. Furthermore, the discrete model converges to the continuous model at a rate proportional to the inverse of the number of products.

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. Suppose that for every element of $x \in x$ there exists a β_i in B for which x is a utility maximizing choice in x.

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 $S^{(n)}$. Let $B^{(n)} \subseteq B$ be the set of taste coefficients that make $C(n)$ a maximizing choice from the set $S^{(n)}$.

Theorem 13. *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. As $n \to \infty$ we first demonstrate that as $\lim_{n\to\infty} C(n) = x^*$, where $x^* = \arg\max_{x\in X}\ u(x, p(x), \beta_i)$ is the utility maximizing choice from all of X when the taste coefficients are β_i. This follows immediately from the fact that the utility function is strictly concave. Since $D_{\beta,x}u(x, p(x), \beta_i)$ has full rank, there is a unique β_i that makes x^* the utility maximizing choice. Let $B^* = \cap B^{(n)}$. Let $\{\beta^{(n)}\}$ be any sequence with $\beta^{(n)} \in B^{(n)}$. Since $\lim_{n\to\infty} C(n) = x^*$ it follows from the continuity of the utility function that for and $\varepsilon > 0$ there exists a sufficiently large n such that if $x = C(n)$, $u(x, p(x), \beta_i^{(n)}) \leq u(x^*, p(x), \beta_i^{(n)}) + \varepsilon$. By assumption 2, there is a unique inversion in the limit from utility maximizing choices to taste parameters. Therefore, $\lim_{n\to\infty} \beta_i^{(n)} = \beta_i$. □

In addition to establishing that in the limit the preference parameters can be uniquely recovered, we can also establish a rate of convergence. Let A_{ij} be defined, as in the text. Obviously, the $\{A_{ij}\}_{j=1}^J$ form a partition of B. Let m denote the Lebesgue measure, it follows immediately that:

$$\begin{aligned} \Sigma_{j=1}^J m(A_{ij}) &= m(B) \\ \frac{\Sigma_{j=1}^J m(A_{ij})}{J} &= \frac{m(B)}{J} \end{aligned} \tag{80}$$

Since the set B is bounded, it must be the case that $\frac{m(B)}{J} \to 0$ which in turn implies that the average Lebesgue measure of A_{ij} converges to zero at a rate proportional to $\frac{1}{J}$.

9 References

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

Albert, J. and S. Chib (1993), "Bayesian Analysis of Binary and Polychotomous Data," *Journal of the American Statistical Association,* 88, 669-679.

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.

Bartik, T. J. (1987), "The Estimation of Demand Parameters in Hedonic Price Models," *Journal of Political Economy*, **95:1**, 81-88.

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.

Blundell, R., M. Browning, and I. Crawford (2001), "Nonparametric Engel Curves and Revealed Preference," IFS Working Paper W197/14.

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

Brown, J. N., and H. S. Rosen (1982), "On the Estimation of Structural Hedonic Price Models," *Econometrica,* 50:3.

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.

Epstein, L. G. (1982), “Integrability of Incomplete Systems of Demand Functions,” *The Review of Economic Studies*, 49:3, 411-425.

Epstein, L. G. (1983), “Erratum: Integrability of Incomplete Systems of Demand Functions,” *The Review of Economic Studies*, 50:3, 581.

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

Geweke, J. (1996), ”Monte Carlo Simulation and Numerical Integration,” in H. Amman, D. Kendrick and J. Rust (eds.), *Handbook of Computational Economics.* Amsterdam: North-Holland, 731-800.

Geweke, J. (1997), ”Posterior Simulators in Econometrics”, in D. Kreps and K.F. Wallis (eds.), *Advances in Economics and Econometrics: Theory and Applications, vol. III.* Cambridge: Cambridge University Press, 128-165.

Geweke, J., M. Keane, and D. Runkle (1994), “Alternative Computational Approaches to Statistical Inference in the Multinomial Probit Model,” *Review of Economics and Statistics,* 76, 609-632.

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.

Imbens, G. and W. K. Newey (2001), “Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity,” Working Paper.

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

Manski, C. F. (1995), *Identification Problems in the Social Sciences*, Cambridge, Mass: Harvard University Press.

Manski, C. F. (1997), "Monotone Treatment Response," *Econometrica*, 65:6, 1311-1334.

Manski, C. F., and J. Pepper (2000), "Monotone Instrumental Variables: With an Application to the Returns to Schooling," *Econometrica*, 68:4, 997-1010.

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

Mas-Colell, A. (1977), "The Recoverability of Consumer Preferences from Market Demand Behavior," *Econometrica*, Vol. 45, No. 6. (Sep., 1977), pp.1409-1430.

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

McCulloch, R., and P. Rossi (1996) "Bayesian Analysis of the Multinomial Probit Model," working paper, Graduate School of Business, University of Chicago.

McFadden, D. and M. K. Richter (????), "Stochastic Rationality and Revealed Stochastic Preference," in J. Chipman, D. McFadden, and M.K. Richter (eds.), *Preferences, Uncertainty, and Optimality, Essays in Honor of Leo Hurwicz*, Westview Press: Boulder, CO, 161-186.

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.