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

Patrick Bajari[†]
Stanford University and NBER

C. Lanier Benkard[‡]
Stanford University and NBER

February 11, 2003

Abstract

We study the identification and estimation of Gorman-Lancaster style hedonic models of demand for differentiated products for the case when one product characteristic is not observed. Our identification and estimation strategy is a two-step approach in the spirit of Rosen (1974). Relative to Rosen's approach, we generalize the first stage estimation to allow for a single dimensional unobserved product characteristic, and also allow the hedonic pricing function to have a general, non-additive structure. In the second stage, if the product space is continuous and the functional form of utility is known then there exists an inversion between the consumer's choices and her preference parameters. This inversion can be used to recover the distribution of random coefficients nonparametrically. For the more common case when the set of products is finite, we use the revealed preference conditions from the hedonic model to develop a Gibbs sampling estimator for the distribution of random coefficients. We apply our methods to estimating personal computer demand.

*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 three anonymous referees, and 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 support. Any remaining errors are our own.

[†]Dept. of Economics, Stanford University, Stanford, CA 94305

[‡]Graduate School of Business, Stanford University, Stanford, CA 94305-5015

1 Introduction

This paper considers the identification and estimation of hedonic models (Gorman (1980), Lancaster(1966, 1971)) of consumer demand in differentiated products markets. The application of hedonic models was pioneered by Rosen (1974). However, Rosen's approach, while widely used in the past in some empirical literatures (e.g., housing and labor markets), has been neglected in many other empirical literatures (e.g., I.O.), and has otherwise drawn some recent criticisms (e.g., Brown and Rosen (1982), Epple (1987), Bartik (1987), and others) that have proven difficult to address.

We believe that there are two main reasons that hedonic models have not been used in empirical work more widely. The first is that the model as outlined by Rosen (1974) assumes perfect competition and a continuum of products. While these assumptions may be appropriate in markets such as housing, in many other markets they are not. In I.O. applications, for example, imperfect (oligopolistic) competition is often specifically a topic of interest. Imperfect competition has several implications to hedonic models. One is that in imperfectly competitive markets it is less likely that the hedonic price function would have an additively separable form. Another is that is rare for oligopolistic markets to contain more than a few hundred products, making the continuous product space assumption unpalatable. Thus, in an effort toward making the hedonic model more widely applicable, in this paper we relax these two assumptions: we allow the price function to take on a general nonseparable form, and we develop estimators for consumers' preference parameters for the case when there are a small (finite) number of products in the market.

Another important reason that hedonic models have not been more widely used is the strict assumption that all characteristics are perfectly observed. In practice, it is common for this assumption to lead to revealed preference violations. For example, it is not uncommon for data sets to contain two products with positive demand in the same period, where one of the products is "better" in every dimension of characteristics space, and also has a lower price. In such cases, there is no set of parameters under which the hedonic model can rationalize the observed demands. And, it is unclear how to proceed, other than to use an alternative model such as the standard discrete choice econometric models. Thus, in this paper we also relax the assumption of perfect observability of all characteristics, and instead allow one product characteristic to be observed by the consumer but not by the econometrician (see also Berry and Pakes (2001) for an alternative approach to this problem). We believe that, with these three generalizations (nonseparable price function, discrete product space, unobserved characteristics), the hedonic model will be substantially easier to apply to standard data sets, a claim that we investigate further in section 6 by estimating a demand system for personal computers.

Similarly to Rosen's approach, our estimation procedure has two stages. In our first stage, price data are used to recover both the hedonic pricing function and the unobserved characteristics. Generalizing the first stage to allow for nonseparability and unobserved product characteristics requires some additional assumptions, and we consider four specific cases of interest. The first case we consider is when the unobserved product characteristics are independent of the observed product characteristics (using the results of Matzkin (2002)). This first case is a stronger version of the (mean) independence assumption commonly used in the empirical literature. In the second case, the consumer maximizes utility by first choosing a "model" and then choosing an "options package." This second case is a nonparametric analog to fixed effects in a linear model. Many product markets, such as automobiles and computers, have this feature. The third case is when there is at least one market in which prices are not a function of the observed characteristics. Our leading example for this case is packaged goods industries, in which different varieties of the same brand frequently have the same price, while different brands have different prices. The fourth case is a nonseparable nonparametric instrumental variables approach due to Imbens and Newey (2001).

In the second stage of the estimation procedure we show that demand data, at either the aggregate or household level, can be used to recover consumers' preference parameters. The problem of identification of preferences is well understood for the case where consumers' entire demand functions are known (see Mas-Colell (1977)). This case corresponds to a data set containing many observations for each consumer or household under different pricing regimes.

However, in our experience, data sets are seldom this rich and frequently only contain a handful or even a single observation per consumer. Therefore, we consider two other cases. First, if the choice set is continuous, then the household level preference parameters must satisfy a set of first order conditions that require the marginal rate of substitution between a continuous product characteristic and the composite commodity to equal the implicit price of that product characteristic. If the functional form of the utility function is known and the parameter vector is of equal or lesser dimension than the characteristics vector, then these first order conditions can be used to recover household level random coefficients. By aggregating household level random coefficients, the population distribution of random coefficients can be nonparametrically identified.

We believe that it is more common in empirical applications for the product space to be finite (i.e., for the market to contain a small number of products) so we go on to consider this second case. In this case, an individual consumer's random coefficients typically are not identified from the revealed preference conditions even if the parametric form of utility is known. Instead, the revealed preference conditions imply that each individual's taste coefficients lie in a set. We show that this set tends to be smaller when there are more products in the market, eventually converging to a singleton. We show how to use these sets for each individual to construct bounds on the population distribution of random coefficients. The procedure is shown to converge to the population distribution of taste coefficients as the number of products becomes large. We also develop a computationally simple Gibbs sampling procedure that can be used to estimate the population distribution of taste coefficients when the product space is finite.

Our estimation procedure avoids the criticisms of Brown and Rosen (1982), Epple (1987) and Bartik (1987) primarily by being less ambitious (though more general) than the Rosen paper with respect to the second stage estimation of preferences. Rosen's approach attempts to obtain higher order approximations to the utility function by imposing homogeneity across individuals, and runs into an identification problem in the process. We instead retain all of the heterogeneity across individuals (by allowing each individual to have different utility parameters), with the cost being that we are less ambitious about the extent to which we can learn the true functional form of utility. We also show in section 4.2 that if individual demographic data is available and some additional independence assumptions hold, then homogeneity can be used to obtain better approximations to utility. Note that Ekeland, Heckman, and Nesheim (2001) makes significantly more progress than we do in this regard.

Our model of demand has much in common with the models used in the recent empirical literature in I.O. on applications of demand estimation beginning with Berry (1994) and Berry, Levinsohn, and Pakes (1995) [BLP], and including Nevo (2001), Petrin (2002), and many others. In particular, our use of the unobserved product characteristics in the demand model is drawn from that literature. The primary difference between our demand model and the BLP models is the fact that our model does not explicitly allow for an *iid* random error term in the utility function.[1] Another important difference is that our approach facilitates nonparametric identification and estimation of the individual preference parameters. Our identification/estimation approach also differs substantially from that in the BLP literature in that we estimate the unobserved product characteristics using the price function whereas BLP models estimate them using the aggregate demand function.[2]

Relative to other approaches in the literature, our approach has some benefits and some costs. One benefit is that our approach is general to a wide variety of equilibrium assumptions, both dynamic and static. We also allow for a general functional form for the price function, and we place no restrictions on the joint distribution of the random taste coefficients. Allowing for rich heterogeneity in consumer preferences has been shown to be important in applications on target marketing (e.g. McCulloch and Rossi(1996)), segregation (e.g. Bajari and Kahn (2002)), and would also be beneficial in any application addressing distributional questions. Another potential benefit of our approach is the fact that our model has no random error terms in the utility function (see Berry and Pakes (2001) or Bajari and Benkard (2002) for a

[1] Note, however, that expanding the second stage of our model to allow for a random error term is a straightforward extension.

[2] We thank an anonymous referee for pointing out that our first stage estimation could be replaced with a first stage that instead used demand functions and nonparametric IV. Such an approach would be the nonparametric analog to BLP. However, the details of nonparametric IV have not, to our knowledge, been fully worked out yet, and nonparametric IV estimation is likely to be significantly more difficult on several dimensions than our approach.

discussion of this issue).

On the other hand, the additional generality of our approach also comes at a cost. Our first stage estimation is likely to require significantly more data than existing alternatives such as BLP or other discrete choice econometric models. Because we allow for nonseparability, we also make a stronger independence assumption. We also allow for only a single-dimensional, vertically differentiated unobserved characteristic.[3] Additionally, the lack of supply side information in our model could be seen as a drawback to the model in that supply side information has the ability to add greater efficiency to the estimation. Finally, there may be cases where it would be desirable to have a random error in the utility function. Note that, while we have not addressed either of these last two issues in this paper, we believe that it would be straightforward to incorporate them into our framework.

The rest of the paper proceeds as follows. Section 2 introduces the model and notation, and proves that if demand is given by the hedonic model then there exists an equilibrium price function. Section 3 shows identification of the price function and unobserved product characteristics (first stage). Section 4 shows identification of preferences (second stage). Section 5 presents econometric estimators consistent with the arguments of sections 3 and 5. Finally, section 6 applies the estimators to estimating personal computer demand.

2 The Model

In our model, a product $j \in \mathcal{J}$ is a finite dimensional vector of characteristics, (x_j, ξ_j), where $x_j = (x_{j1}, ..., x_{jK})$ is a K dimensional vector of characteristics observed by both the consumer and the econometrician, and ξ_j is a scalar that represents a characteristic of the product that is observed only by the consumer. The set $X = \cup_{j \in \mathcal{J}} (x_j, \xi_j) \subseteq R^{K+1}$ represents all products that are available to consumers in the market.

[3] Goettler and Shachar (2001) relaxes both of these assumptions in the BLP framework. Benkard and Bajari (2002) uses techniques similar to those of this paper to recover a multidimensional unobserved characteristic in the context of price indexes.

Let p_{jt} denote the price of product j in market $t \in T$. The elements of T can be thought of as markets separated by space or time. 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 given by $u_i(x_j, \xi_j, c) : X \times R_+ \to R$. The price of the composite commodity is normalized to one. Consumers have income y_i and consumer i's budget set in market t, $B(y_i, t)$, must satisfy:

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

Consumer i in market t solves the maximization problem,

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

2.1 The Price Function

This section shows under weak conditions that, in any equilibrium, the model above implies that prices in each market must have the following properties: (i) there is one price for each bundle of characteristics, (ii) the price surface is increasing in the unobserved characteristic, and (iii) the price surface satisfies a Lipschitz condition. The theorem relies only on consumer maximization, the fact that prices are taken as given, and some simple assumptions on consumers' utility functions. Most importantly, it is independent of supply side assumptions.

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 unobserved product characteristic. However, without A3 the price function is not guaranteed to be increasing in ξ.

Theorem 1. *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.

The equilibrium price function for market t is denoted $\mathbf{p}_t(x_j, \xi_j)$. It is a map from the set of product characteristics to prices that satisfies $p_{jt} = \mathbf{p}_t(x_j, \xi_j)$ 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.

2.2 Discussion

Because the theorem above is based on demand side arguments only, these results are general to many types of equilibria, both dynamic and static. Note, however, that the theorem only speaks to the prices of products actually observed with positive demand. A consequence is that for some cost functions and some demand patterns, certain bundles may never be observed. For example, this is likely to be the case if the cost function was discontinuous. In such cases it also seems likely that there would be a selection problem in the price function estimation.

Before discussing the identification and estimation of the model, we feel that some further explanation of the price function is necessary. The price function in each market is an equilibrium function that is dependent upon market primitives. It does not tell what the price of a good would be if that good is not already available in the market. If a new good were added, in general all the prices of all the goods would change to a new equilibrium, and thus the whole price function would change as well. The price function would also change if any other market primitives were to change, such as consumer preferences, marginal production costs, or if a good already in the market were to be produced by another multi-product firm. This is the primary reason for the fact that we have to treat the price function as being possibly 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. Therefore, it can be used to define the consumer's budget set.

As to what functional form the price function may take on, even some very simple models of competition would suggest that the equilibrium price function may 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 all the characteristics. Thus, we feel it would not be appropriate for us to assume that the price function was additively separable in the unobserved product characteristics and proceed using standard econometric approaches such as OLS or IV. Instead, we proceed by maintaining the general form above.

3 Identification of the Price Function and the Unobserved Characteristics

3.1 Identification Using Independence.

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

First, consider identification of the price surface if prices are observed without error. We begin with two assumptions.

A4 ξ is independent of x.

A5 For all markets t and all x, $\mathbf{p}_t(x,\cdot)$ is strictly increasing, with $\frac{\partial \mathbf{p}_t(x,\xi)}{\partial \xi} > \delta$ for all (x,ξ) for all t and some $\delta > 0$.[4]

Assumption A4, which requires full independence, is a strengthening of the mean independence assumption commonly used in the literature. The stronger independence assumption is required because we allow the model to be nonseparable. There are many examples of markets in which we would expect independence to be violated. For example, if either marginal costs are nonseparable in all characteristics, or demand is nonseparable in all characteristics (e.g. characteristics are complements), then independence is likely to be violated. The implications of proceeding even if the assumption is violated are discussed in section 5.4.

If independence holds, then the support of the unobserved product characteristics does not depend on the observed characteristics so that $\mathbf{p}_t : A \times E \to R$, where $A \subseteq R^K$ is the support of x, and $E \subseteq R$ is the support of ξ. Assumption A5 follows from more primitive assumptions on consumer preferences (see Theorem 1).

For the case where there is a single market, that is $T = \{1\}$, and no measurement error in prices, Matzkin (2002) shows under weak conditions that both the functional form of $\mathbf{p}_1(\cdot)$ and the distribution of the unobserved product characteristics, $\{\xi_j\}$, are identified up to a normalization on ξ. The first part of our identification proof follows Matzkin (2002), the only differences being that her results are extended to cover the case of many markets, and we use an alternative normalization that facilitates estimation.

[4] 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. The main theorem only requires $\delta \geq 0$. The proof with measurement error requires $\delta > 0$.

Let I be the set of price functions satisfying A5,

$$I = \{\mathbf{p}' : A \times E \to \mathbb{R} \mid \text{for all } x \in X,\ \mathbf{p}'(x, \cdot) \text{ is strictly increasing}\} \tag{2}$$

Since the unobserved product attribute has no inherent units, it is only possible to identify it up to a monotonic transformation. Thus, without loss of generality, we assume that a normalization has been made to ξ such that the marginal distribution of ξ is $U[0,1]$. Technically, this amounts to normalizing ξ using its distribution function. The practical impact of this normalization is discussed below in section 4.5.

We define identification to be identification within the set satisfying the normalization made above,

Definition 1. *The function* $\mathbf{p}$ *is* ***identified in*** I *if*

i. $\mathbf{p} \in I$*, and*

ii. For all $\mathbf{p}' \in I$*,*

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

We now show that identification holds in the case where prices are measured without error.

Theorem 2. *If prices are observed without error and A4-A5 hold, then* p_t *is identified in* I *for all* t*. Furthermore,* $\{\xi_j\}$ *is identified.*

Proof. We first show how to construct the unobserved product characteristics using the conditional distribution of prices,

$$\begin{aligned}
F_{p_t|x=x_j}(p_{jt}) &= Pr(\mathbf{p_t}(x,\xi) \leq p_{jt}|x = x_j) \\
&= Pr(\xi \leq \mathbf{p_t}^{-1}(x, p_{jt})|x = x_j) \\
&= Pr(\xi \leq \mathbf{p_t}^{-1}(x_j, p_{jt})) \\
&= \mathbf{p_t}^{-1}(x_j, p_{jt}) \\
&= \xi_j
\end{aligned}$$

To construct the price function for each market we need only invert the above relationship,

$$\mathbf{p_t}(x_0, e_0) = F^{-1}_{p_t|x=x_0}(e_0) \tag{3}$$

□

From the proof of the theorem we can see that, in the absence of measurement error, identification of the unobserved product characteristics can be obtained in a single cross-section. Therefore, for example, identification is obtained even if products are observed in only one market. We mention this because it implies that identification is obtained even if the unobserved product characteristics for each product change over time, as is commonly assumed in the empirical literature. It also means that identification is obtained even if the distribution of the unobserved product characteristics changes over time (or across markets).

In the appendix we show that cross-market variation can be used to obtain identification when prices are measured with error. However, in that case each product must be observed in many markets.

3.2 Identification Using "Options Packages"

This section provides an alternative set of assumptions that also provide identification and that we believe may be satisfied in some applications. In some markets, consumers 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 (RAM, processor speed/type, hard drive,...). 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 unobserved product characteristics.

Let z denote a model and $\mathcal{Z}$ denote the set of all models. The set of models induces a partition of $\mathcal{J}$. The map $\pi : \mathcal{J} \to \mathcal{Z}$ associates products (j) with models (z). The inverse image of z under π is the set of products that are model z, where each product has a possibly different options package x. The model z is observable and x and z have joint distribution $F_{x,z} : A \times \mathcal{Z} \to R$.

The first assumption in this section says that ξ is shared by products that are the same model:

A6. 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 also need there to be a "baseline" or standard options package that is available for all models z. We formalize this requirement using the following assumption,

A7. There exists an $\bar{x} \in A$ such that for all $z \in \mathcal{Z}$, $f(\bar{x}|z) > 0$.

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 normalize ξ such that $F_{\xi|x=\bar{x}}$ is $U[0,1]$. The next theorem shows identification for the case where prices are observed without error.

Theorem 3. *If prices are observed without error and A5-A7 hold, then p_t is identified in I for all t. Furthermore, $\{\xi_j\}$ is identified.*

Proof. For each product j, let j^* be a product such that $\pi(j) = \pi(j^*)$ and $x_{j^*} = \bar{x}$. Such a product exists for every model $\pi(j)$ by A7. Then, similarly to the previous section,

$$\xi_j = F_{p_t|x=\bar{x}}(p_{j^*t})$$

This equation identifies $\{\xi_j\}$.

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

$$\mathbf{p_t}(x_0, e_0) = p_{kt} \text{ for } k \in \mathcal{J} \text{ such that } \xi_k = e_0 \text{ and } x_k = x_0 \tag{4}$$

$\square$

Again in this case, identification of the unobserved product characteristics is obtained in a single cross-section. However, unlike the independence case above, in this case identification can be obtained in a single cross section even if prices are measured with error (see appendix). The reason for the difference is that we now observe many products in each market that are known to have the same value of ξ. A consequence of this is that it is not necessary to observe products in more than one market to obtain identification with measurement error in the prices.

Another consequence of observing many products in each market that are assumed to have the same value for the unobservable is that the model is overidentified, and is therefore testable. If there is no measurement error in prices, then the model is rejectable in the sense that assumption A6 may be violated in the data.

3.3 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. 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. We do not assume that the researcher knows which market this is.

A8 There exists a market, t, such that $\mathbf{p_t}(x, \xi) = f(\xi)$, with $f_\xi > 0$.

In our opinion, assumption A8 is not likely to hold in the majority of applications, but may hold in some specialized circumstances. A8 is most likely to hold in markets where "quality" is the primary differentiating feature of the product with respect to determining price. For example, in many packaged goods markets, even though consumers may have strong preferences over "flavors", which would typically be observable as dimensions of x, all flavors of a given product line often have exactly the same price, while different product lines have different prices.[5]

Second, we also need weak monotonicity of prices in all of the characteristics,

A9 For all markets t, $\mathbf{p_t}(x, \xi)$ is weakly increasing in all of the observed characteristics, x, and strictly increasing in the unobserved characteristic, ξ.

We think that A9 is likely to hold in many applications. If all individuals have monotone preferences over all characteristics, then A9 holds by an argument similar to that of Theorem 1. However, A9 might hold even if this were not the case. For example, if marginal costs were sufficiently increasing in all characteristics, then A9 would also hold.

Theorem 4. *If prices are observed without error, A8 and A9 hold, and (x, ξ) have full support on $A \times E$, then p_t is identified in I for all t. Furthermore, $\{\xi_j\}$ is identified.*

[5] Specifically, suppose that in one market the price function is $\mathbf{p_t}(x, \xi) = w \cdot x + \xi$ where x is a vector of "flavor dummies" and all of the elements of w are the same.

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 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:

$$\mathcal{J}' = \{j' \in \mathcal{J} \ : \ (x_{j',1}, ..., x_{j',k}) = x^*, \text{ and } p_{j',t} \leq p_t(x) \text{ for all } t\} \tag{5}$$

It follows from A8 and A9 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 $\{\xi_j\}$. A normalization thus identifies the $\{\xi_j\}$ and $F_{x,\xi}$. Identification of $p(x, \xi)$ follows directly. □

Note that the proof above requires the fact that all products are observed in many markets.

3.4 Identification Using Instruments

In the event that the unobserved product characteristics are not independent of all of the observed characteristics, a fourth possible approach for identifying and estimating the first stage of the model would be to use nonseparable nonparametric instrumental variables. The details of such an approach have not to our knowledge been worked out in general. However, Imbens and Newey (2001) provide estimators for triangular systems that would likely be applicable to our model in many applications.

The primary difficulty with using the Imbens and Newey instrumental variables approach is that it necessitates finding instruments that determine (in the sense of functional dependence) the value of the "endogenous" observed characteristic(s) but that are independent of the unobserved characteristic. In the kinds of applications that we are interested in, the past empirical literature has relied on independence assumptions between the observed and unobserved characteristics, showing that such instruments may be difficult to find. However, there are some applications where it is possible. For example, as instruments for the endogenous characteristics (e.g. racial make-up) of a given neighborhood, Bayer, McMillan, and Ruben (2002) use the fixed characteristics (e.g. housing stock characteristics) of housing in surrounding neighborhoods.

4 Identification of Preferences

The results of Section 3 provide sufficient conditions for identification of $\mathbf{p}_t(x, \xi)$ and $\{\xi_j\}$. If the consumer demand function is known for all $\mathbf{p} \in I$, then the results of Mas-Colell (1977) provide sufficient conditions for identification of consumer i's weak preference relation, $\succeq_i$.[6] Unfortunately, most data sets are not this rich. Therefore, we consider the identification of consumer preferences in cases where less information about the demand function is available. The first case we consider is when the choice set is continuous, but there are a finite number of observations per individual. This case is similar to that of the Rosen (1974). We also consider the more common case where the set of products is discrete in section 4.3.

4.1 Continuous Choice Set and a Finite Number of Observations Per Individual

Typically only a small number of choices are observed per consumer, often just one. In that case, recovery of the entire weak preference relation is not possible. Figure 1 graphically illustrates the consumer's utility maximization problem. The problem is projected into two dimensions so that the consumer maximizes utility over just two product characteristics. By standard arguments, at the chosen bundle, the marginal rate of substitution between these characteristics is equal to the slope of the consumer's budget set. This information provides only local information about preferences at each chosen bundle.

Under some additional assumptions, the range in which the indifference curve must lie can be

[6] In Mas-Colell (1977), the budget sets are linear. Therefore, it is sufficient to know the consumer demand function for all $\mathbf{p} \in I$ that are linear in (x, ξ) to apply these results.

bounded. If consumers have diminishing marginal rates of substitution, 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 preferences are monotonic, then the indifference curve must lie everywhere below the indifference curve for Leontief preferences. Together, the two assumptions allow us to conclude that the indifference curve must lie in the shaded area of Figure 1. 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.[7]

A simple way to narrow down the range of possibilities is to place parametric restrictions on the consumer's indifference curves. Many discrete choice models in the literature assume that utility is linear or log-linear in (x, ξ, c), e.g.,

$$u_{ij} = \beta_{i,1} \log(x_{1,j}) + \cdots + \beta_{i,K} \log(x_{K,j}) + \beta_{i,\xi} \log(\xi_j) + c. \tag{6}$$

In the equation above, the utility of household i for product j depends on household specific preference parameters, $\beta_i = (\beta_{i,1}, ..., \beta_{i,K}, \beta_{i,\xi})$. If there is an interior maximum, then the first order conditions for utility maximization are

$$\frac{\beta_{i,k}}{x_{k,j}} = \frac{\partial \mathbf{p}_t}{\partial x_{j,k}} \qquad \text{for } k = 1, ..., K \tag{7}$$

$$\frac{\beta_{i,\xi}}{\xi_j} = \frac{\partial \mathbf{p}_t}{\partial \xi_j}. \tag{8}$$

These first order conditions can be solved simply for the unknown preference parameters of the individual,

$$\beta_{i,k} = x_{k,j} \frac{\partial \mathbf{p}_t}{\partial x_{j,k}} \qquad \text{for } k = 1, ..., K \tag{9}$$

$$\beta_{i,\xi} = \xi_j \frac{\partial \mathbf{p}_t}{\partial \xi_j}. \tag{10}$$

If the price function, $\mathbf{p}_t$, and unobserved characteristics $\{\xi_j\}$ are known, then in this example

[7] 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.

household i's preference parameters, $\beta_i = (\beta_{i,1}, ..., \beta_{i,K}, \beta_{i,\xi})$, can be recovered even if only a single choice of the household, (x_j, ξ_j), is observed. By aggregating the decisions of all of the household in market t, $F_t(\beta)$, the population distribution of taste coefficients in market t can be learned.

In general, we characterize an agent by a B dimensional parameter vector $\beta_i \in R^B$. Since the previous section has shown that the unobservables, $\{\xi_j\}$, are identified by the price function, we proceed as if ξ is known and write the utility function as

$$u_i(x, c) = u(x, y_i - \mathbf{p}(x); \beta_i). \tag{11}$$

where the dependence of utility on ξ is dropped to simplify notation.

Agents choose the element $x \in X$ that maximizes utility. If both u and $\mathbf{p}(x)$ are differentiable, then the first order necessary conditions are

$$\frac{\partial}{\partial x_k}\{u(x, y_i - \mathbf{p}(x); \beta_i)\} = 0 \text{ for } k = 1, ..., K \tag{12}$$

Let $x(\beta)$ denote the optimal choice of x conditional on β. The first order conditions can be implicitly differentiated to yield

$$\begin{aligned} x'(\beta) &= -\left[D_{x,x}\overline{u}\right]^{-1} D_{x,\beta_i}\overline{u} && (13) \\ \text{where } \overline{u}(x; \beta) &= u(x, y_i - \mathbf{p}(x); \beta_i) && (14) \end{aligned}$$

Theorem 5. *Suppose $\beta_i \in \mathcal{B} \subseteq R^B$, where $\mathcal{B}$ is an open convex subset and $x \in R^K$. Then if $x'(\beta)$ is locally negative definite or positive definite, then β_i is locally identified. If $K = B$, and $x'(\beta)$ is globally positive definite or negative definite, then $x(\beta)$ is one-to-one.*

Proof: The first part of the theorem follows from the local version of the inverse function theorem. The second part follows from the global inverse function theorem since if (13) is everywhere positive or negative definite, then $x(\beta)$ is one-to-one so that the preferences are globally identified. (see Gale and Nikaido (1965)). Q.E.D.

Theorem 5 places tight restrictions on the types of utility functions that can be identified using the choice data. Conditional on knowing the price surface $\mathbf{p}$, we can identify at most K random coefficients per choice observation. While this may seem like a negative result, we do not view it that way. Even just a first order approximation to the utility function may be good enough for many applications. For example, the experiment of removing a single good from the market to evaluate the consumer surplus obtained from the good (e.g. Petrin (2002)) would involve only local changes to utility if the choice set is rich. Additionally, if more than one choice per household is available, the first order conditions can be used to provide higher order approximations to the utility function.

We also note that while the theorem may guarantee that preference parameters are identified for many nonlinear functional forms for utility, in many cases solving for the preference parameters implied by consumer's choice may be a nontrivial numerical problem.

4.2 Imposing Homogeneity

In this section we investigate the potential for using demographic data and homogeneity assumptions in order to obtain better approximations to the utility function even with only a single observation per consumer. Suppose that a set of demographic variables, w_i, are observed for each consumer i. Suppose further that consumers preference parameters are a function of the demographic variables and a taste shifting vector, η_i, where

$$\beta_{ik} = f_k(w_i) + \eta_{ik} \tag{15}$$

and η_i is mean independent of the demographic variables. Then it is straightforward to use standard techniques to estimate the functions f_k.

Equation (15) uses covariation in tastes across individuals with different demographics to identify systematic changes in preferences with demographic variables, while retaining the heterogeneity in tastes across individuals. Another way to interpret (15) is that homogeneity restrictions can be used to obtain a more flexible approximation to the utility function.

With this interpretation, our approach is similar in spirit to that of Blundell, Browning, and Crawford (2003), as well as Ekeland, Heckman, and Nesheim (2001). Bajari and Kahn (2002) implements this procedure on an application to racial segregation in cities.

4.3 Discrete Product Space

In practice, there are at least three reasons why the continuous choice model might not provide a good approximation to choice behavior. First, the number of products in the choice set may not be sufficiently large that the choice set is approximately continuous. Second, many product characteristics are fundamentally discrete (e.g., "power steering", "leather seats"). Third, some consumers may choose products at the boundaries (e.g., the fastest computer).

In place of the marginal conditions in (12), when the product space is discrete, consumer maximization implies a set of inequality constraints. If consumer i chooses product $j \in \{1, ..., J\}$ then

$$u(x_j, \xi_j, y_i - \mathbf{p}(x_j, \xi_j); \beta_i) \geq u(x_k, \xi_k, y_i - \mathbf{p}(x_k, \xi_k); \beta_i) \text{ for all } k \neq j. \tag{16}$$

Therefore, it must be that $\beta_i \in A_{ij}$, where

$$A_{ij} = \{\beta_i : \beta_i \text{ satisfies } (16)\}. \tag{17}$$

If the choice set is finite, the A_{ij} sets will typically not be singletons, implying that the parameters β_i are not identified. However, that does not mean that the data is non-informative. If the choice set is rich, the A_{ij} sets may be small. In the appendix it is shown that if all of the characteristics are continuous and the choice set is compact, then as the number of products increases, the A_j sets converge to the individual taste coefficients β_i. In applications where the A_{ij} sets are large enough that the lack of identification matters, we show below that it is possible to proceed in two ways. First, the A_{ij} sets can be used 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.

4.4 Non-Purchasers and Outside Goods

In our model, individuals that choose not to purchase any product are handled similarly to those that do purchase. The decision not to purchase any product is the same as the consumer spending all of her income on the composite commodity c. That is, it is as if she purchases a bundle that provides zero units of every characteristic and carries a zero price. In either the continuous or discrete product space cases, this would imply a set of inequalities for nonpurchasers of the form,

$$u(0,0,y_i);\beta_i) \geq u(x_k,\xi_k,y_i - \mathbf{p}(x_k,\xi_k);\beta_i) \text{ for all } k. \tag{18}$$

These inequalities could then be used similarly to those above in (16) to locate nonpurchasers' preference parameters. Note that (18) provides only inequalities and therefore there is an identification problem for non-purchasers even if the product space is continuous.

4.5 The Choice of Normalization for ξ

Above we demonstrated that when there are only a small number of observations per individual, the individual's complete preference ordering is not identified. We also showed how functional form assumptions on the utility function can be used to overcome this. Clearly, the choice of functional form for utility has some influence on the results obtained. In this section, we note that the choice of normalization for ξ from the first stage interacts with the functional form assumption on utility such that it too may influence the results.[8] For example, in figure 1, if one of the dimensions shown represented the unobserved characteristic, ξ, it would be possible using a monotonic transformation to renormalize ξ such that the budget set was non-convex. We conclude that, just as we should take care in choosing the functional form for utility, it is necessary to take care in choosing the first stage normalization. It may also make sense to check the results for sensitivity to this normalization.

There have been several normalizations proposed in the past econometrics literature. Each

[8] We thank an anonymous referee for pointing out that due to this feature of the model it is possible to construct examples under which the results obtained would be quite misleading.

has a different economic interpretation. The $U[0,1]$ normalization used in section 3 above amounts to scaling ξ by its distribution function. Essentially this normalization says that, holding x constant, changing ξ by 0.01 leads to a 1% move upward in the price distribution. Unfortunately, this does not tell us anything about the implications of the normalization for the budget set, and furthermore it is hard to come up with any economic justification for this normalization.

A second normalization has been proposed by Matzkin (2002). Matzkin (2002) normalizes ξ such that at some bundle $\bar{x}$, ξ equals the price of the product. This normalization has a more straightforward economic interpretation. It says that at $\bar{x}$ increasing ξ by one increases price by a dollar. However, without also knowing the marginal effect of x on price, we again cannot say for sure what the implications of this normalization are for the budget set. If x has an increasing marginal price, as is common in empirical applications, then since ξ has an approximately constant marginal price (for x near $\bar{x}$) the budget set should be convex in ξ.

We do know that there are a variety of reasonable cases under which the budget set would be convex in ξ. For example, this would be true if the cost of producing characteristics were sufficiently increasing, or if characteristics are sufficiently complementary in production. It is also true in the additively separable case, again provided that the cost of producing characteristics is sufficiently increasing, or if there are complementarities. In our experience with data on observed characteristics, it is usually the case that budget sets are convex in observed characteristics. These arguments suggest that the budget set is also likely to be convex in the unobserved characteristic in many applications. Therefore, one possible approach to the normalization would be to choose a normalization of ξ such that the budget set is convex ex post.

Note that while the normalization may change the results somewhat, the effect of the normalization is mitigated in estimation by the estimated preference parameters for ξ. Furthermore, if enough observations are available for each individual, then preferences over ξ are identified and the normalization of ξ does not affect the results. In our application (section 6), with just aggregate data, we found that this mitigating effect was strong and the demand system estimates were not very sensitive to the particular normalization of ξ used.

5 Estimation

5.1 Estimation, Stage 1: 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 3.1. In particular, we assume that x and ξ are 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.[9]

Let $\hat{F}_{p_t|x=x_0}(e_0)$ be an estimator for the conditional distribution of prices given $x = x_0$ at the point e_0 in market t. For example, a kernel estimator (such as those outlined in Matzkin (2002)) or a series estimator (such as those outlined in Imbens and Newey (2001)) could be used. In section 6 we found that a local linear kernel estimator (Fan and Gijbels (1996)) worked best. Define an estimator for ξ by the following,

$$\hat{\xi}_{jt} = \hat{F}_{p_t|x=x_j}(p_{jt})) \tag{19}$$

While Matzkin (2002) does not explicitly consider estimation of the unobservable, the asymptotic properties of the estimator in (19) 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. However, after plugging in the estimated true prices,

[9] A previous draft of this paper contained estimators for the measurement error case. Please contact the authors for details.

the asymptotic properties of the estimator would change.[10]

5.2 Estimation of Preferences, Continuous Product Space

Next, a strategy is outlined for estimating preferences in the case of one observation per individual and a simple functional form for utility. When multiple observations per individual are available, other, more flexible specifications, can be estimated similarly.

To illustrate the approach, assume that the consumer's utility takes the form in equation (6). Then the first order conditions imply that equations (9) and (10) must hold. This suggests the following estimator for β_i

$$\widehat{\beta}_{i,k} = x_{k,j}\frac{\partial \widehat{\mathbf{p}}_t}{\partial x_{j,k}} \text{ for } k = 1, ..., K \tag{20}$$

$$\widehat{\beta}_{i,\xi} = \widehat{\xi}_j \frac{\partial \widehat{\mathbf{p}}_t}{\partial \xi_j} \tag{21}$$

where $(x^i_j, \hat{\xi}^i_j)$ represents the (estimated) bundle chosen by individual i and $\frac{\partial \widehat{\mathbf{p}}_t}{\partial x_{j,k}}$ 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 β_i. 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 (19) above (except using (3)) and using either a kernel or series-based approach. Matzkin (2002) also provides a direct estimator for the price function derivatives.

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. 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

[10] This is because the measurement error estimator would have dimension $K + 1$ while the estimator $\hat{F}$ has dimension K.

taste coefficients and demographics in the population, as well as to estimate preferences as a function of demographics as outlined in section 4.2.

5.3 Estimation of Preferences, Discrete Product Space

In this section, we propose an approach to estimating β_i when the product space is discrete. Section 4.3 demonstrated that the taste coefficients are typically not identified in this case. The strategy, therefore, is to recover the *sets* of taste coefficients that are consistent with consumers' choices. This approach is in the spirit of the bounds literature (see Manski (1995, 1997) and Manski and Pepper (2000)). We also borrow heavily from the literature on Bayesian estimation of discrete choice models (Albert and Chib (1993), Geweke, Keane, and Runkle (1994), and McCulloch and Rossi (1996)).

To illustrate our approach, suppose that only data from a single market is used. In that case, the bounds estimator of β_i is A_{ij}. The problem with estimating these bounds is that, when there are a large number of products and product characteristics, the A_{ij} sets are high dimensional and determined by a large number of inequalities, making it difficult to characterize them analytically. Instead, we propose to use numerical methods. We cast the problem of estimating the taste coefficients into a Bayesian paradigm. Specifically, we construct a likelihood function and a prior distribution over the parameters such that the support of the posterior distribution corresponds to the A_{ij} sets. We then derive a simple Gibbs sampling algorithm to simulate from this 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 inequalities (16).

The inequalities (16) generate a likelihood function in a natural fashion. The likelihood that a consumer with taste coefficients β_i chooses product j is:

$$L(j|x, y_i, \beta_i) = \begin{cases} 1 & \text{if } u(x_j, y_i - p_j; \beta_i) \geq u(x_k, y_i - p_k; \beta_i) \text{ for all } k \neq j \\ 0 & 0 \text{ otherwise} \end{cases}. \tag{22}$$

That is, consumer i chooses product j so long as her taste coefficients imply that product

j is utility maximizing. For technical convenience, the prior distribution for β_i, $p(\beta_i)$ is a uniform distribution over the region $\mathcal{B}$. 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

$$p(\beta_i|C(i), x, p) \propto \pi(\beta_i)L(j|x, \beta_i). \tag{23}$$

The posterior distribution is uniform over those $\beta_i \in \mathcal{B}$ that are consistent with the agents choice. So long as $\mathcal{B}$ completely covers all of the A_{ij} sets, the posterior is uniform over A_{ij}.

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:

$$\int g(\beta_i)p(\beta_i|C(i), x, p) \tag{24}$$

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:

$$\lim_{S\to\infty} \frac{1}{s}\sum_{s=1}^{S} g(\beta_i^{(s)}) = \int g(\beta_i)p(\beta_i|C(i), x, p) \tag{25}$$

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 (1996, 1997).

Suppose that household i chooses product j. The first step in developing a Gibbs sampler is to use equation (23) to find the distributions,

$$p(\beta_{i,1}|x, p, C(i) = j, \beta_{i,-1}) \tag{26}$$

$$p(\beta_{i,2}|x, p, C(i) = j, \beta_{i,-2}) \tag{27}$$

$$\vdots \tag{28}$$

$$p(\beta_{i,K}|x, p, C(i) = j, \beta_{i,-K}). \tag{29}$$

If the specification of utility is linear in the β_i and X_j, it is straightforward to derive the conditional densities (26).[11] For example, in the model (6), if j is utility maximizing for household i

$$\sum_l \beta_{i,l} \log(x_{l,j}) + y_i - p_j \geq \sum_l \beta_{i,l} \log(x_{l,k}) + y_i - p_k \text{ for all } k \neq j, \tag{30}$$

which implies that:

$$\beta_{i,1} \geq \frac{\sum_{l \neq 1} \beta_{i,l}(\log(x_{l,k}) - \log(x_{l,j})) + (y_i - p_j) - (y_i - p_k)}{\log(x_{1,j}) - \log(x_{1,k})} \text{ if } x_{1,j} > x_{1,k} \tag{31}$$

$$\beta_{i,1} \leq \frac{\sum_{l \neq 1} \beta_{i,l}(\log(x_{l,k}) - \log(x_{l,j})) + (y_i - p_j) - (y_i - p_k)}{\log(x_{1,j}) - \log(x_{1,k})} \text{ if } x_{1,j} < x_{1,k}. \tag{32}$$

Since both prior distribution and the likelihood are uniform, it follows that the conditional distribution (26) is uniform on the interval $[\beta_{1,\min}, \beta_{1,\max}]$, where

$$\beta_{1,\min} = \max\left\{ \min_{\beta_1|\beta_{-1}} B, \max\left\{ \frac{\sum_{l \neq 1} \beta_{i,l}(\log(x_{l,k}) - \log(x_{l,j})) + (y_i - p_j) - (y_i - p_k)}{\log(x_{1,j}) - \log(x_{1,k})} \text{ such that } x_{1,j} > x_{1,k} \right\}\right\} \tag{33}$$

$$\beta_{1,\max} = \min\left\{ \max_{\beta_1|\beta_{-1}} B, \min\left\{ \frac{\sum_{l \neq 1} \beta_{i,l}(\log(x_{l,k}) - \log(x_{l,j})) + (y_i - p_j) - (y_i - p_k)}{\log(x_{1,j}) - \log(x_{1,k})} \text{ such that } x_{1,j} < x_{1,k} \right\}\right\}. \tag{34}$$

The conditional distribution for the remaining β's is also a uniform distribution defined by inequalities that are analogous to (33) and (34). So long as β_i enters the utility function linearly, the conditional distributions are uniform.

Let $\beta_i^{(0)} = (\beta_{i,1}^{(0)}, \beta_{i,2}^{(0)})$ be an arbitrary point in the support of the posterior. The Gibbs sampling algorithm proceeds as follows:

[11] If the support of the posterior distribution is not connected, Gibbs sampling is not guaranteed to converge. However, if the β_i enter into the utility function linearly, as in equation (6), it can easily be shown that the sufficient conditions for convergence described in Geweke (1994) are satisfied.

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.

This algorithm is quite simple to program since at each step it only requires the econometrician to compute upper and lower bounds similar to (33) and (34) and draw a sequence of uniform random numbers. The sequence of random draws obtained can be used to construct bounds on the distribution function.

An alternative to the bounds approach would be to construct a point estimate for the distribution of tastes for the entire population of consumers. Let $F(\beta_1, ..., \beta_K)$ be the cumulative distribution function for the K taste coefficients. It follows that:

$$\begin{aligned} F(\overline{\beta}_1, ..., \overline{\beta}_K) &= \Pr(\beta_1 \leq \overline{\beta}_1, ..., \beta_K \leq \overline{\beta}_K) \\ &= \lim_{S \to \infty} \frac{1}{S} \sum_{s=1}^{S} 1\{\beta_1 \leq \overline{\beta}_1, ..., \beta_K \leq \overline{\beta}_K\}. \end{aligned} \tag{35}$$

The sample analog of the last expression, using the Gibbs draws, can be used as an estimator for F. This estimator uses the uniform prior to choose one of many possible distributions consistent with the data.

The algorithm can also be used to estimate more general models of choice. Suppose, for example, that consumers are observed more than once. If household i is observed to choose n_i times, then $n_i(J-1)$ inequalities are implied by maximization. The conditional distributions used in Gibbs sampling can then be derived analogously to (33) and (34). The algorithm can also be extended to the cover estimation error in the unobserved product characteristics by using an estimate of ξ_j and proceeding as above. If ξ_j is estimated imprecisely, so that it has a non-degenerate distribution $F(\xi_j)$, the posterior can be simulated by first drawing $\xi_j \sim F(\xi_j)$ for $j = 1, ..., J$ and then, for each draw, using the Gibbs sampling algorithm above.

Perhaps the greatest difficulty in implementing this algorithm is finding an initial value $\beta_i^{(0)}$ that satisfies the set of inequalities (16). There are several solutions to this problem. The simplest one is to use the first order conditions from a continuous choice model to find such a point. So long as the estimated budget set is everywhere convex, this approach is simple and guaranteed to work. More generally, finding a starting value for β_i can be cast as a linear programming problem that can be solved using a standard numerical package.

A second difficulty in practice is when there are no parameter values that satisfy the inequalities in (16) (so that A_{ij} is empty for some i and j). The fact that the first stage estimator assigns higher values of ξ to products with higher prices makes this unlikely to occur. However, we speculate that it may be possible for 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 and would suggest generalizing the utility function.

In the appendix it is shown that the sets A_{ij} shrink to β_i at a rate of $\frac{1}{J}$. Therefore, when the number of products is large, the identification of the preference parameters β_i is similar to inverting the first order conditions as in the previous subsection.

5.4 Results When Independence is Violated

In this section we discuss the implications to the results if the estimators above are applied when the independence assumption is violated. The implications to the first stage estimation are similar to what they would be if running OLS when mean independence is violated, except that the argument only holds locally due to the nonparametric approach. If ξ is (locally) positively correlated with one of the x variables, then in the first stage part of the price effect of ξ would be falsely attributed to that x, such that the marginal price of x was biased upward and the marginal price of ξ was biased downward. In the second stage, this bias would lead to similar biases in our estimates of consumer's preference parameters: the preference for x would be biased upward, and the preference for ξ biased downward. This can most easily be seen by looking at equations (20) and (21).

Clearly this bias would influence marginal willingness-to-pay estimates for characteristics (i.e., the taste distributions). What is less clear are the implications for marginal willingness-to-pay for goods (e.g. the elasticity of demand). Since goods represent bundles of both observed and unobserved characteristics, and since both marginal willingness-to-pay and marginal prices are biased, to some extent these biases offset each other in the demand system for goods. We have verified this intuition analytically by looking at some simple models. In some special cases (such as linear utility and linear price function), the biases exactly cancel each other. However, this is not true generally. We conclude that if the independence assumption is violated then the demand system estimators are likely to be biased. In general, the sign and magnitude of the bias depends on the shape of the equilibrium price function and individual preferences in the market being studied.

6 Application to Computer Demand

In this section we apply our proposed estimators to demand for desktop personal computers. Our data comes from the *PC Data Retail Hardware Monthly Report* and includes quantity sold, average sales price, and a long list of machine characteristics for desktop computers. Please see Benkard and Bajari (2002) for a more detailed discussion of this data.[12]

The raw data set contained 29 months of data, but in this paper we use only the data for the last period, December 1999, covering 695 machines. We chose to use data from a single period both to keep the estimation simple and also to test how well our methods work on a single cross-section, a case where traditional demand estimation techniques do not typically work well. The data set reportedly covers approximately 75% of U.S. retail computer sales. The raw data contained a large number of characteristics, including dummies for each individual processor type. We eliminated the processor dummies in favor of a CPU benchmark

[12] This paper is available at http://www.stanford.edu/~lanierb/.

variable.[13] The final data set contained 19 characteristics, including five operating system dummies (Win 3.1, NT 4.0, NT, Win 98, Win 95) plus CPU benchmark, MMX, RAM, hard drive capacity, SCSI, CDROM, DVD, modem, modem speed, NIC, monitor dummy, monitor size (if monitor supplied), zip drive, desktop (versus tower), and refurbished. Summary statistics for this data are given in table 1.

6.1 First Stage Results: Unobserved Product Characteristics

Because it would be difficult to use nonparametric techniques on a 19-dimensional system, in the application we concentrate on the three most important continuous characteristics: CPU benchmark, RAM, and hard drive capacity. However, we also found that the results were not as clean if the remaining 16 characteristics were simply omitted. Therefore, for the purposes of the first stage regressions, we made the assumption that the price function is additively separable in the remaining characteristics, with coefficients that are known and equal to the values obtained from an OLS regression. Thus, the first stage estimates are obtained using price data after the linear effect of the other 16 characteristics has been removed (OLS coefficients listed in table 1).

While this assumption was largely made for convenience in the estimation, we believe that additive separability is likely to hold for many of the remaining characteristics because many of them represent such things as peripheral devices that can easily be removed from the computer and bought or sold on a secondary market. Therefore, arbitrage should limit deviations of the price function from additive separability in these characteristics.

Our first stage estimates were obtained using a local linear kernel estimator (Fan and Gijbels (1996)) corresponding to equation (19), with a normal kernel. Equation (19) was derived using the $U[0,1]$ normalization and thus results in estimates of ξ that are limited to the unit interval. Bandwidths were chosen by eye to be large enough that small wiggles in the distribution function were eliminated. This results in a bandwidth that is likely to be oversmoothing the

[13] CPU benchmark was obtained from www.cpuscorecard.com. A regression of CPU benchmark on chip speed interacted with chip dummies yielded an R^2 of 0.999.

distribution function somewhat relative to the MSE minimizing bandwidth. For our purposes it is useful to oversmooth since we use estimates of the price function derivatives to solve for starting values for the Gibbs sampling algorithm in the second stage of estimation. Note, however, that because we are treating the product space as finite, we only need the first stage estimates of ξ in order to estimate the demand system. We do not need estimates of the price function except for the fact that they provide convenient starting values for the Gibbs estimator.

An important issue with respect to our approach is how precisely it is possible to estimate the ξ's for each product. Table 2 shows the distribution of standard errors for the estimated ξ's using both an asymptotic approximation (from Fan and Gijbels (1996)) and also a bootstrap that assumes *iid* sampling of products (an assumption that we recognize is not likely to reflect the economics of the product entry and exit processes, but that was maintained purely to provide a check for the asymptotic standard errors). The asymptotic standard errors average 0.01, with 95% of the estimated ξ's having standard errors less than 0.018. The bootstrap standard errors were generally very similar but slightly larger in magnitude.[14]

In summary, we found that, with very few exceptions, it was possible to estimate the unobserved product characteristics quite precisely. We believe that this result reflects several features of our model and data. One important feature of the model is the fact that in our model ξ corresponds to a distribution function (or a quantile), and is therefore much easier to estimate than, for example, the corresponding density. Another reason for the precision of the results is that in our data the characteristics space is quite densely filled so that, with the exception of a few outlying products, every product has close neighbors in characteristics space.

We also estimated the bias in the ξ estimates, and found it to be small, but more importantly we found that the rank ordering of the ξ estimates was nearly identical whether or not the estimated bias term was included. Hence, we left the estimated bias out of our final first

[14] 508 of 695 products had larger estimated standard errors with the bootstrap than with the asymptotic formula.

stage estimates.

6.2 Second Stage Results: Preferences and Demand Curves

In estimating the demand system, high dimensional systems are easily handled. Thus, for the purposes of the demand system estimation, we included all 19 observed characteristics as well as the unobserved characteristic. For the purposes of this example, we used the quasi-linear functional form for utility (as in equation (6)), linear in the logs of the continuous characteristics (CPU benchmark, hard drive, RAM, ξ) and linear in the remaining characteristics, including price. The unobserved characteristics obtained from the first stage were normalized such that budget sets were always convex. We also tried several alternative normalizations and found that the effects of the choice of normalization on the results were small.

We used the Gibbs sampling estimator described in section 5.3 to estimate the distribution of the individual taste coefficients. Starting values for the Gibbs algorithm were obtained by using the values of the taste coefficients obtained if it were assumed that the true product space was continuous (as in section 5.2).[15] After a series of initial draws that were discarded, we simulated 1000 taste coefficient draws per product, for a total of 695000 (20-dimensional) draws.[16] The draws were then re-weighted by the observed total demand for each product in order to make calculations from the posterior distribution of the taste coefficients. The approach of simulating a fixed number of draws per product, as opposed to varying the number of draws simulated with demand for the product, guarantees similar precision of the simulation results across the entire taste distribution space.

Figures 2-6 graph the (kernel smoothed) taste distributions for the three continuous characteristics, as well as SCSI and the unobserved characteristic. In all cases the taste distributions were converted into marginal willingness-to-pay figures so that the absolute value of the coefficients could be interpreted as dollars. Except for CPU benchmark, the willingness-to-pay

[15] In some cases these starting values were not inside the appropriate A_{ij} set for the discrete case and hence we used a numerical search algorithm to find a point nearby that was inside the correct set.

[16] Including the initial draws, this process took about an hour on our Sun Ultra 60 workstation.

distributions generally look somewhat log-normally distributed, and seem reasonable given 1999 prices. Modal marginal willingness-to-pay for RAM is between one and two dollars per megabyte, while modal marginal willingness-to-pay for hard drive capacity is approximately 1.25 cents per megabyte.

One result that surprised us was that the data appear to be quite informative even for the dummy characteristics such as SCSI. For dummy characteristics, the revealed preference bounds obtained for each individual are generally looser than they are for continuous characteristics. In our data, two things appear to be true. First, the individual bounds are more informative than we expected. While it is not generally possible to place very tight bounds on the SCSI taste coefficient independently of the other coefficients, extreme values of the SCSI coefficient often require extreme values of the other coefficients in order to rationalize the individual's choice. Therefore, the revealed preference sets have relatively small volume in areas with extreme values for the SCSI coefficient, which causes our estimator to place small weight on extreme values even at the individual level. The second thing is that the individual posterior distributions turn out to be quite informative about the shape of the overall population distribution once they have been aggregated up.

Table 3 shows the correlation in tastes across this subset of characteristics. We find that there is a high positive correlation between the taste parameters for most of the characteristics. Notably, this is true even for the unobserved characteristic. While this last result makes intuitive sense, it is reassuring that the taste distributions reflect this correlation despite the fact that unobserved characteristics were assumed to be independent of the observed characteristics in the first stage estimation.

Using the estimated taste distributions it is easy to simulate estimates of the demand curves and price elasticities for each product using explicit aggregation. However, when we did this we found that the implied demand elasticities were counterintuitively high (the median own price elasticity was around -100). After further investigation, we attribute this finding to the perfect information assumption made in the model. The model assumes that each consumer costlessly obtains perfect information regarding all 695 products in the choice set. With such a large number of products in the choice set, products are on average quite close in price and characteristics space. According to the model, products are sufficiently similar and sufficiently close in price that if consumers had perfect information then the price elasticities should be very high.

However, such high price elasticities do not seem to us to be consistent with profit maximization and the likely levels of entry and fixed costs. We interpret the result instead as a rejection of the perfect information assumption. We believe that in actuality consumers have search costs and therefore, when making a purchase, they only acquire information about a small number of products, perhaps only a handful.

In order to try to capture the effect of imperfect information, we reestimated the model using only those products that are most prominent and that most consumers could easily obtain information on. The easiest way to do this was to eliminate products that had small market shares from the choice set. Thus, we reestimated the model eliminating any product that sold less than 5000 units, corresponding to a minimum market share of 0.75%. This left the 24 largest products, together accounting for 72% of total sales.[17] After reestimating the demand model including only these products, the estimated demand elasticities ranged from -4 to -72, with a median elasticity of -11, results which we think are much more plausible.

Table 4 describes the top five products in terms of sales and table 5 lists the estimated cross-price elasticities for these five products. Two features of the elasticities are worth mentioning. The first is that the products that are most similar substitute the most highly (e.g., the two Compaq's). The second is that there are some cross price elasticities that are exactly zero, a feature that can be generated by the hedonic model but cannot be generated by standard econometric discrete choice models. For example, in the table the HP6535 and the Compaq5461 do not substitute at all. Presumably this is because there are other products located between these two products in characteristics space.

[17] A drawback of using this product selection method is that it tended to select quite similar products of the type that typically sold the most units, as opposed to picking the best selling product in each class. Thus, we believe that the price elasticities estimated may still be biased upward relative to a model in which the search process was modeled explicitly.

Figure 7 shows the estimated demand curve for the HP6535, with the actual price and quantity denoted on the graph. The estimated demand curves in our model must be downward sloping by utility maximization. However, because the shape of the curve is determined by individuals' marginal willingness-to-pay for each product, and we have placed no restrictions on the taste distributions, any shape of downward sloping demand may result from the estimation. For example, it would be easy to obtain kinked demand curves. Despite that, the shape of the demand curve obtained for the HP6535 is typical of those we obtained for other products. We generally find that demand curves for computers are quite steep for high prices, with a flattening in the middle near the observed price, and then steeper again for low prices. The upper steep portion is obtained because typically there is some small mass of consumers who strongly prefer this particular product. The flattening in the middle reflects a large mass of consumers that are willing to switch to other neighboring products without very much compensation. The steepening at low prices reflects the fact that at low prices the product in question has already captured a large market share and thus there are few consumers left to gain by lowering price further. One difference that showed up for some products is that when there is a close substitute present with slightly higher price the demand curve may head more quickly to the vertical axis, without showing the steep portion in figure 7.

While standard approaches to estimating demand systems rely heavily on cross-market variation in order to identify price elasticities, the revealed preference conditions in the hedonic model allow identification of the shape of the whole demand curve even from a single cross section of data. In order to emphasize this feature of the model, in this example we used data from just a single market so there was no cross-market information at all. In our opinion, the shape of the demand curve in figure 7 is quite reasonable compared with the demand curves implied by standard discrete choice econometric models, which necessarily asymptote to both axes, and compared with standard multi-stage models, which are typically quadratic. These results suggest that the revealed preference conditions from the hedonic model may be a good alternative way of obtaining identification of demand systems.

Finally, we wish to note that the results reported in this section represent essentially all of the estimations we ran for this model and were thus obtained without the need for any kind of specification search other than the reduction of the product space noted above.

7 Conclusion

This paper has investigated the identification and estimation of hedonic discrete choice models of differentiated products. Specifically, we showed how to generalize Rosen's (1974) in three primary ways: (1) we allow for one product characteristic to be unobserved by the econometrician, (2) we generalize the first stage estimation to the nonseparable case, (3) we allow for a discrete product space and discrete characteristics in estimating preferences. Our hope is that these generalizations will make it easier to apply hedonic demand models in empirical work in a wider set of applications in the future.

In order to obtain identification of the first stage, we have shown that one of four possible sets of assumptions can be used. Three of the four methods represent nonseparable nonparametric analogs standard linear methods: OLS, fixed effects, and IV. The fourth is a special case in which the quality ordering is uniquely determined by one special market. There may also be other approaches to estimating the first stage. An important source of additional identification that we have left open for future work is the use of conditions obtained from the supply side of the model.

Lastly, by imposing less homogeneity across individuals we believe that we have avoided recent criticisms of the hedonics literature, with the tradeoff being that we obtain identification of preferences primarily through functional form restrictions on the utility function.

A Identification with Measurement Error in Prices

A.1 Independence Case

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

$$y_{jt} = p_{jt} + \epsilon_{jt} \equiv \mathbf{p_t}(x_j, \xi_j) + \epsilon_{jt} \tag{36}$$

We assume classical measurement error:

B1 ϵ_{jt} is *iid*, and $E[\epsilon|x, \xi] = 0$.

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.

For identification with measurement error it is necessary for products to be observed in many markets, as can be seen in the proof of the following theorem.

Theorem 6. *If prices are observed with error, A4, A5, and B1 hold, then p_t is identified in I for all t. Furthermore, $\{\xi_j\}$ is identified.*

Proof. Let

$$\bar{\mathbf{p}}^{\mathbf{T}}(x, \xi) = \frac{1}{T} \sum_{t=1}^{T} \mathbf{p_t}(x, \xi) \tag{37}$$

and let $\bar{\mathbf{p}}_j^{\mathbf{T}} \equiv \bar{\mathbf{p}}^{\mathbf{T}}(x_j, \xi_j)$. For each product we can observe $\bar{\mathbf{p}}_j^{\mathbf{T}}$ 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.

For each product, j, define the set

$$\mathcal{J}_j = \{k \in \mathcal{J} |\ x_k = x_j \text{ and } \lim_{T \to \infty} \bar{\mathbf{p}}_j^{\mathbf{T}} - \bar{\mathbf{p}}_k^{\mathbf{T}} = 0\} \tag{38}$$

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 $\mathcal{J}_j$:

$$p_{jt} = E[y_{kt} | k \in \mathcal{J}_j] \tag{39}$$

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 follows by Theorem 2.

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

A.2 Options Packages Case

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

Theorem 7. *If prices are observed with error, A5-A8 and B1 hold, then p_t is identified in I for all t. 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

$$\mathbf{p_t}(x_j, \xi_j) = E[y_{kt} | k \in \mathcal{J}_j], \tag{40}$$

where the measurement error again averages to zero. The rest of the proof is by Theorem 3. □

B Convergence of the Discrete Model to the Continuous Model

In the text, it was shown that when the set of products is discrete, β_i is not identified. However, as the number of choices become sufficiently large, β_i can learned in the limit. Furthermore, the sets A_j shrink at a rate $\frac{1}{J}$.

To simplify notation, attention is restricted to the case where all product characteristics are observable to both the consumer and the econometrician. Consumer i's utility is written as $u_{ij} = u(x_j, p_j, \beta_i)$. Also, suppose that $\mathbf{p}(x)$ that maps characteristics into prices for any product j. To simplify notation, this function is assumed to be independent of the number of products J in the market, although the results could be modified to cover this case. Three assumptions about the product space and the utility are made:

Assumption 1. All of the product characteristics x_j are elements of X an open, bounded and convex set. Also, all of the β_i lie in $\mathcal{B}$, an open, bounded and convex set.

Assumption 2. For any β_i,when the choice set is all of X , the jacobian $x'(\beta)$, as defined in (13) is everywhere positive definite or negative definite.

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 the set of choices available to consumer i. Let $C(n)$ be the utility maximizing choice for consumer i when she can choose from $S^{(n)}$. Let $B^{(n)} \subseteq \mathcal{B}$ be the set of taste coefficients that make $C(n)$ a maximizing choice from the set $S^{(n)}$. Note that as an implication of assumptions 1 and 2, the global inverse function theorem can be applied and $x(\beta)$ is one-to-one.

Theorem 8. *Suppose that Assumptions 1-2 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 $n \to \infty$, $\lim_{n\to\infty} C(n) = x^*$. Let

$B^* = \cap B^{(n)}$. Let $\{\beta^{(n)}\}$ be any sequence with $\beta^{(n)} \in B^{(n)}$. Suppose that $\beta' \neq \beta^*$ is in B^*. Then for all n and all $\widetilde{x}(n) \in S^{(n)}$, $u(\widetilde{x}(n), \mathbf{p}(\widetilde{x}(n)), \beta') \leq u(C(n), \mathbf{p}(C(n)), \beta')$. Letting $n \to \infty$, it follows that for all $x \in X$, $u(x, \mathbf{p}(x), \beta') \leq u(x^*, \mathbf{p}(x^*), \beta')$. But this contracts the fact that $x(\beta)$ is one-to-one. Q.E.D.

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_j be defined, as in the text. Obviously, the $\{A_j\}_{j=1}^{J}$ form a partition of $\mathcal{B}$. Let m denote the Lebesgue measure, it follows immediately that:

$$\begin{aligned} \Sigma_{j=1}^{J} m(A_j) &= m(\mathcal{B}) \\ \frac{\Sigma_{j=1}^{J} m(A_j)}{J} &= \frac{m(\mathcal{B})}{J}. \end{aligned} \tag{41}$$

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

C References

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

Bajari, P. and M. Kahn (2002), "Estimating Housing Demand with an Application to Explaining Racial Segregation in Cities," Working Paper, Stanford University.

Bajari, P. and C. Benkard (2002), "Discrete Choice Models as Structural Models of Demand: Some Economic Implications of Common Approaches" Working Paper, Stanford University.

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

Bayer, P., R. McMillan, and K. Rueben (2002), "An Equilibrium Model of Sorting in an Urban Housing Market: A Study of the Causes and Consequences of Racial Segregation", working paper, Yale University.

Benkard, C. and P. Bajari (2002), "Hedonic Price Indexes with Unobserved Product Characteristics, and Application to PC's" working paper, Graduate School of Business, Stanford University.

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 (2001), "Estimating the Pure Hedonic Discrete Choice Model," Working Paper, Yale University.

Blundell, R., M. Browning, and I. Crawford (2003), "Nonparametric Engel Curves and Revealed Preference," *Econometrica*, 71:1, 205-240.

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.

Ekeland, I., Heckman, J., and L. Nesheim (2002), "Identification and Estimation of Hedonic Models," IFS Working Paper CWP07/02.

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

Fan, J. and I. Gijbels (1996), *Local Polynomial Modeling and Its Applications,* London: Chapman and Hall.

Gale, D. and Nikaido, Hukukane (1965) "The Jacobian Matrix and Global Univalence of Mappings," *Mathematische Annalen* 159 pp81-93.

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 (2001), "Estimating Spatial Competition in the Network Television Industry," *RAND Journal of Economics*, 32:4, 624-656.

Gorman, T. (1980), "A Possible Procedure for Analysing Quality Differentials in the Egg Market," *Review of Economic Studies*, 47:5, 843-856.

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

Lancaster, K. (1966), "A New Approach to Consumer Theory," *The Journal of Political Economy*, 74:2, 132-157.

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

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

Matzkin, R. (2002), "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.

Nevo, A. (2001), "Measuring Market Power in the Ready-To-Eat Cereal Industry," *Econometrica*, 69:2, 307-342.

Petrin, A. (2002), "Quantifying the Benefits of New Products: The Case of the Minivan," *Journal of Political Economy,* 100, 705-729.

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

D Tables and Graphs

Table 1: Summary Statistics

Variable	Mean	S.D.	Min	Max	OLS Coeff
CPU Bench	1354.5	362.3	516	2544	0.836
RAMMB	74.0	35.1	16	256	3.010
HDMB	9276.8	4850.3	2100	40000	0.008
MMX	0.64	0.48	0	1	-56.971
SCSI	0.01	0.08	0	1	310.747
CDROM	0.67	0.47	0	1	26.478
DVD	0.14	0.35	0	1	32.213
NIC	0.36	0.48	0	1	9.481
Monitor?	0.31	0.46	0	1	29.625
Mon.Size	0.75	3.27	0	15	22.822
ZIP	0.05	0.22	0	1	20.440
DT	0.17	0.37	0	1	25.611
Refurb.	0.09	0.28	0	1	-144.314
No Modem	0.55	0.50	0	1	145.169
Win NT 4.0	0.02	0.14	0	1	-106.374
Win NT	0.17	0.37	0	1	22.567
Win 98	0.58	0.49	0	1	-59.590
Win 95	0.16	0.37	0	1	-42.058
Constant	(Win3.1 omitted)				-590.2
R^2					0.79
N					695

Table 2: Distribution of Standard Errors for Estimates of ξ

	Asymptotic	Bootstrap (1000 samples)
Quantiles:		
Min	0.002	0.002
0.30	0.007	0.007
0.50	0.009	0.009
0.95	0.018	0.022
0.99	0.052	0.091
Max	0.104	0.121
Average	0.010	0.012
N	695	695

Table 3: Correlation Matrix of Taste Coefficients for a Subset of Characteristics

	CPU	RAM	HD	SCSI	ξ
CPU	1.000	0.510	0.357	0.694	0.418
RAM	0.510	1.000	0.533	0.511	0.477
HDM	0.357	0.533	1.000	0.527	0.351
SCSI	0.694	0.511	0.527	1.000	0.393
ξ	0.418	0.477	0.351	0.393	1.000

Table 4: Top Five Products in 12/99

Brand/Model	CPU (Benchmark)	RAM	H.D.	Price	Sales
Hewlett Packard Pavilion 6535	Intel Celeron/466MHZ (1281)	64MB	8.4GB	590	71199
Compaq Presario 5441	AMD A6-2/475MHZ (1076)	64MB	8.0GB	540	54449
Compaq Presario 5461	AMD A6-2/500MHZ (1115)	64MB	10.0GB	727	43029
E-Machines eTower 433	Celeron/433 (1167)	32MB	4.3GB	471	40399
Hewlett Packard Pavilion 6545C	Celeron/500 (1398)	64MB	13.0GB	858	34198

Table 5: Matrix of Cross Price Elasticities for Top Five Products

	HP6535	Compaq5441	Compaq5461	E-Machines	HP6545C
HP6535	-4.14	0.12	0.00	0.43	0.28
Compaq5441	0.17	-5.95	2.98	0.73	0.55
Compaq5461	0.00	2.80	-8.00	0.85	0.11
E-Machines	0.61	0.69	0.91	-10.65	0.66
HP6545C	0.70	0.86	0.18	1.02	-4.46

Figure 1: Global Identification of Indifference Curves

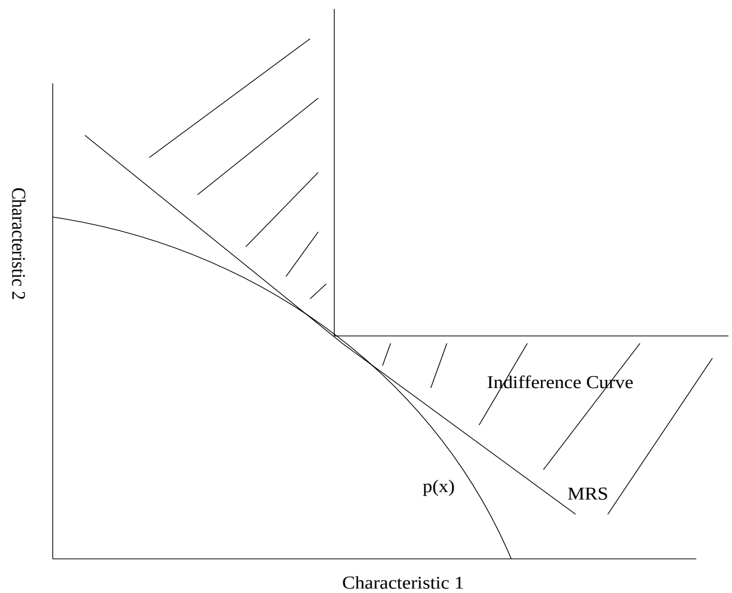

Figure 2: CPU Benchmark Willingness-To-Pay

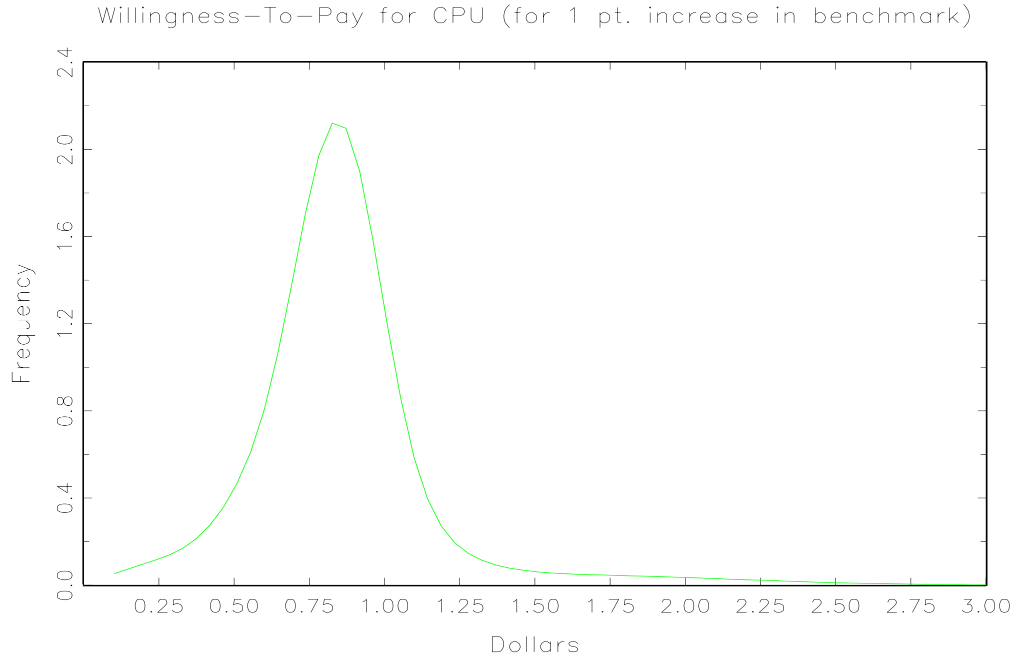

Figure 3: RAM Willingness-To-Pay

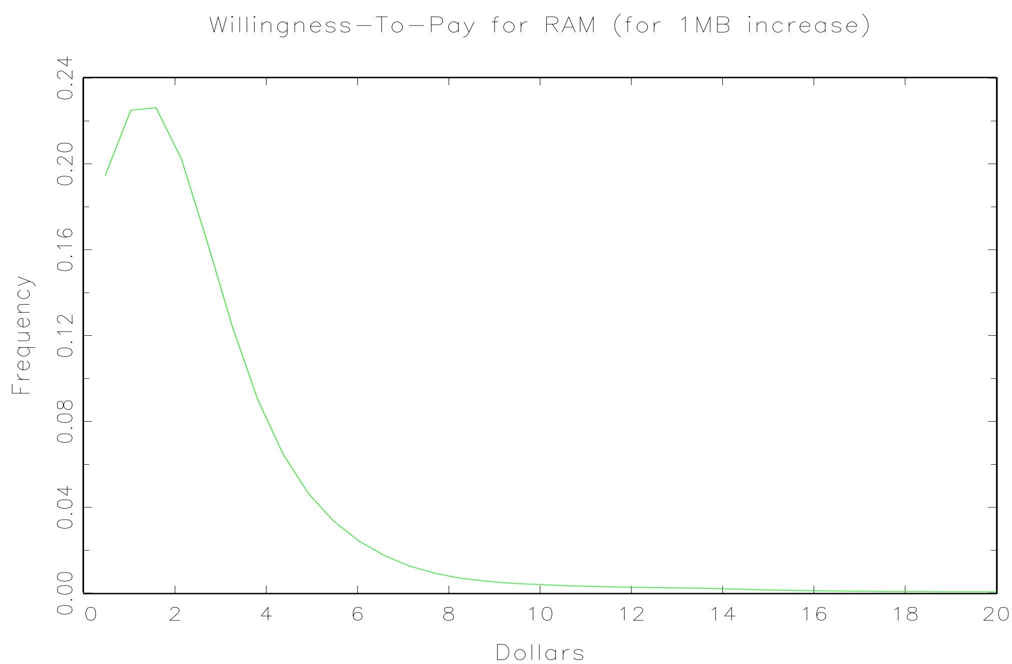

Figure 4: Hard Drive Willingness-To-Pay

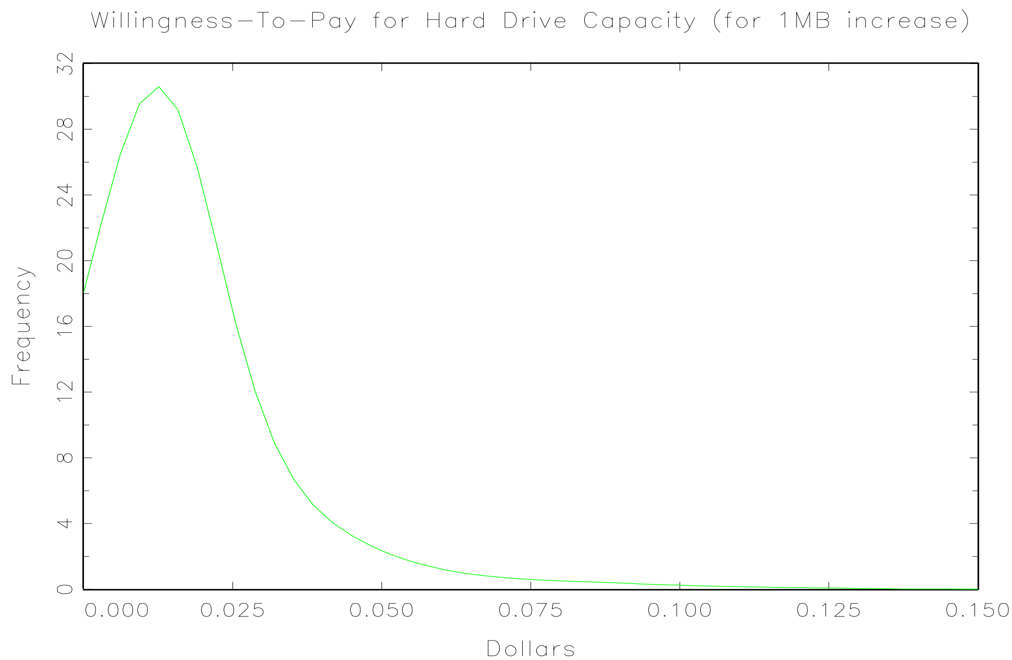

Figure 5: SCSI Willingness-To-Pay

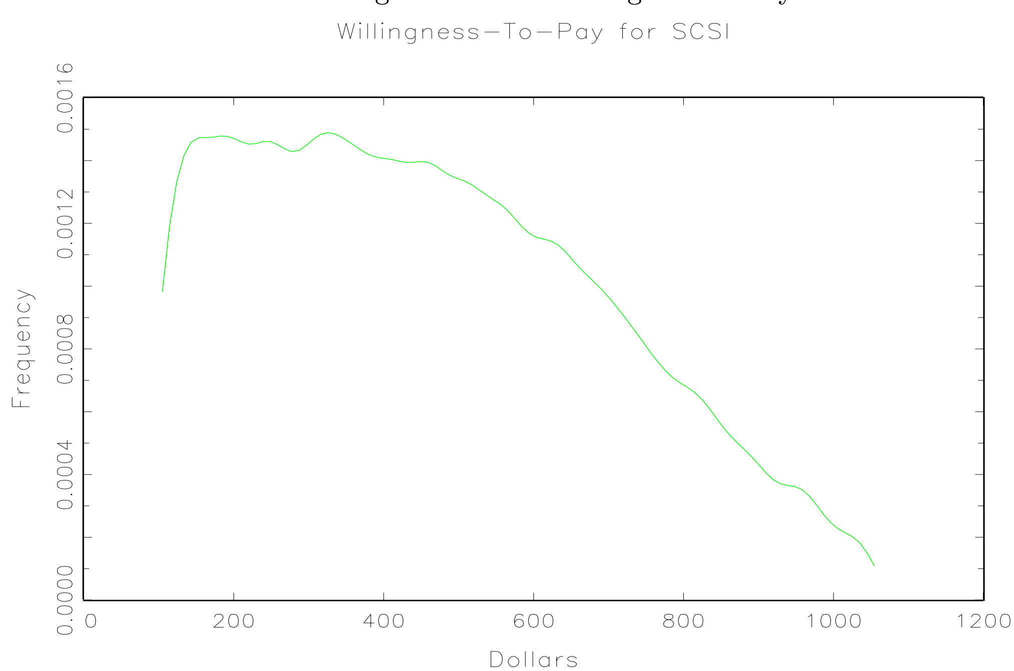

Figure 6: ξ Willingness-To-Pay

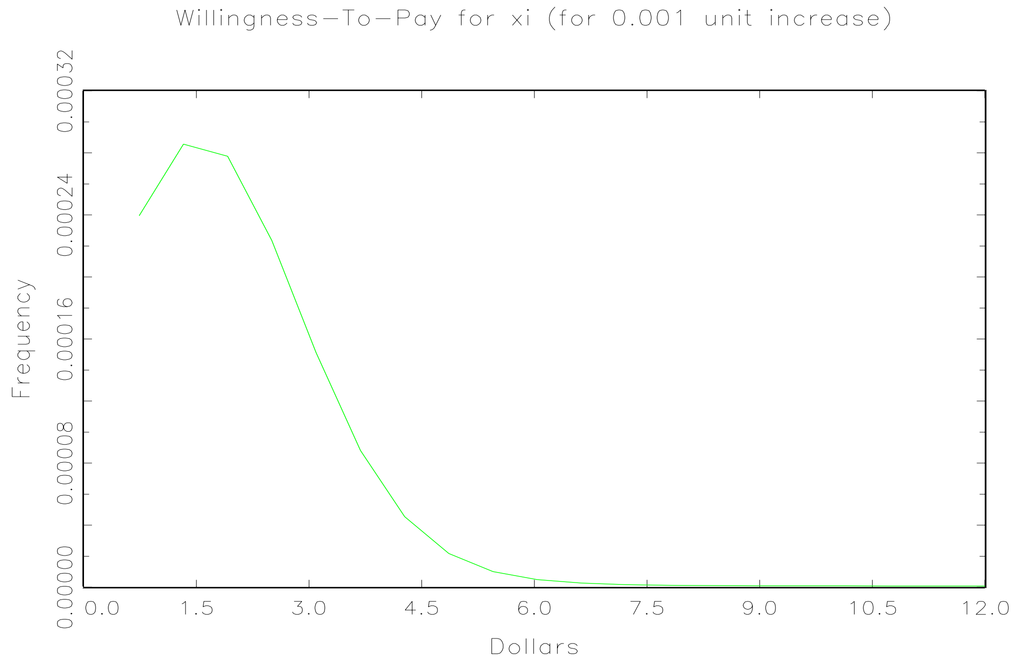

Figure 7: HP6535 Demand Curve