# Demand Estimation with Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach

## Patrick Bajari

*University of Michigan and National Bureau of Economic Research*

## C. Lanier Benkard

*Stanford University and National Bureau of Economic Research*

We reconsider the identification and estimation of Gorman-Lancaster-style hedonic models of demand for differentiated products in the spirit of Sherwin Rosen. We generalize Rosen's first stage to account for product characteristics that are not observed and to allow the hedonic pricing function to have a general nonseparable form. We take an alternative semiparametric approach to Rosen's second stage in which we assume that the parametric form of utility is known, but we place no restrictions on the aggregate distribution of utility parameters. If there are only a small number of products, we show how to construct bounds on individuals' utility parameters, as well as other economic objects such as aggregate demand and consumer surplus. We apply our methods to estimating the demand for personal computers.

## I.   Introduction

This paper reconsiders the identification and estimation of hedonic models (Gorman 1980; Lancaster 1966, 1971) of consumer demand in differentiated product markets. The application of hedonic models was pioneered by Rosen (1974). Rosen's approach has been widely used in the literatures on housing, public economics, environmental economics, and labor markets and somewhat less frequently in those on marketing and industrial organization.[1] Hedonic methods have been applied to problems as diverse as determining the value of clean air, estimating price indexes, and measuring the value of a good reputation in online auctions.

   Rosen's approach to estimating a hedonic demand system consists of two stages. In the first stage, the prices of goods are regressed on the goods' characteristics. The coefficients in this regression are often interpreted as implicit prices or as the consumer's marginal willingness to pay for each characteristic. This type of hedonic pricing regression, which also originated much earlier with Court (1939) and Griliches (1961), remains widely used. Rosen also proposed a second stage that involved regressing the marginal prices of each characteristic at the bundles actually purchased by consumers onto the characteristics of the good and consumer demographic variables. This second stage was intended to recover a demand function for each characteristic. However, it was later discovered (Brown and Rosen 1982; Bartik 1987; Epple 1987) that the second-stage regression had a simultaneity problem because consumers with a high preference for a certain characteristic would naturally purchase bundles that contained large amounts of this characteristic. This simultaneity problem causes inconsistent estimates in the second stage. Epple suggests that this problem can be solved if data are available on many markets in which tastes can be assumed to be the same. However, data of this kind have proved difficult to find, and as a result, Rosen's second stage is not as widely used today.

   This paper revisits this literature with the hope of reviving the use of hedonic models in estimating demand systems. We make several contributions to both stages of the estimation, substantially generalizing the first-stage estimation and taking an alternative approach to the second stage that avoids the criticisms above and that we think is more appealing in practice.

   In the first stage, Rosen assumed that the market being studied had a continuum of products and perfect competition. These assumptions may be appropriate in some markets, such as housing, but they are unreasonable in many others. In industrial organization applications,

---

[1] At present, Rosen (1974) has over 1,250 citations in the *Social Science Citation Index.*

for example, imperfect (oligopolistic) competition is often specifically a topic of interest. It is also rare for oligopolistic markets to contain more than a few hundred products, making the continuous product space assumption unreasonable. Rosen also assumed that all product characteristics are perfectly observed. The industrial organization and marketing literatures have found that unobserved product attributes are important both in explaining the data and in causing endogeneity in prices. One way this can be seen is that it is common for the perfect observability assumption to lead to some products being strictly dominated. For example, it is common to see two products in a market in which one of the products is "better," meaning that it has at least as much random-access memory, central processing unit speed, hard drive, and so forth, and the better product also has a lower price, yet the dominated (inferior) product still has some demand. In such cases it seems likely that there is something about the dominated product that we do not observe that is attractive to some consumers. Moreover, under the assumption that all characteristics are perfectly observed, there is no set of parameters under which the hedonic model can rationalize the observed demands. In our data on the demand for computers, despite the fact that we have 18 observed characteristics, a staggering 357 of the 695 products are dominated in this manner.

We relax all three of these assumptions (perfect competition, continuum of products, and perfect observability of characteristics). First, we show that if demand is given by the hedonic model, then there always exists a function mapping characteristics to prices, regardless of the form of competition and even if there are a small number of products. This discipline comes from the demand side of the model. The existence of the price function justifies running the first-stage hedonic price regressions in a much wider set of applications. In addition, we show that if there is one product characteristic that is not observed and this product characteristic is preferred by every consumer, then the price function must be increasing in the unobserved product characteristic. The fact that the price function is increasing in the unobserved product characteristic will allow us to recover it during the first-stage estimation.

We also generalize Rosen's first stage by considering the identification and estimation of a price function with a priori unknown general form, nonseparable in the unobserved product characteristics. Generalizing the first stage to allow for nonseparability requires some additional assumptions for identification, and we consider three cases that we think represent the situations that are most likely to be encountered in applications. In the first case, the unobserved product characteristics are independent of the observed product characteristics. This first case, which we expect to be the most useful in applications, is a slightly stronger version of the (mean) independence assumption commonly

used in various empirical literatures (industrial organization, marketing, housing, etc.). In the second case, consumers choose both a "model" and an "options package." This second case represents a nonparametric analogue to fixed effects in a linear model. Many product markets, such as automobiles and computers, have this feature. The last case is a nonseparable, nonparametric instrumental variables approach similar to that of Imbens and Newey (2003).

Like Rosen, we also consider a second stage in which the goal is to recover consumers' preferences so that we can compute a demand system, welfare, and other economic objects of interest. However, we take an alternative approach, and our results are somewhat more positive than those in the recent hedonics literature. If there are many products such that the choice set is continuous, then a consumer's product choice must satisfy a set of first-order conditions that require the consumer's marginal rate of substitution between each continuous product characteristic and the composite commodity to equal the implicit (or marginal) price of that product characteristic at the chosen bundle. Clearly, if the consumer is observed only once, a feature of most data sets, then it is not possible to learn the consumer's entire weak preference relation from observing this single choice. Rosen tried to solve this problem by imposing homogeneity across individuals. We instead take a semiparametric approach inspired by the recent empirical literatures in industrial organization and marketing (Berry, Levinsohn, and Pakes [1995], Rossi, McCulloch, and Allenby [1996], Nevo [2001], Petrin [2002], and many others). In order to identify consumers' preferences, we impose structure on the utility function by assuming that it takes a known parametric form. However, we place no restrictions on the aggregate distribution of preferences across consumers. If the parametric form of the utility function is known, then the first-order conditions determining the consumer's optimum can be used to recover household-level random coefficients, household by household. By aggregating household-level random coefficients, we obtain the population distribution of random coefficients nonparametrically.

However, as mentioned above, we think that it is also important to consider what happens in applications in which there are a small number of products (such that the product space is discrete) since that is the case in many applications. In that 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 the individual's taste coefficients lie in a set. This set tends to be smaller when there are more products in the market, eventually converging to a singleton if all the characteristics are continuous. In the spirit of pioneering work by Manski and Pepper (2000), Manski and Tamer (2002), and Haile

and Tamer (2003), we show how these sets for each individual can be used 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 show how to construct bounds for other economic objects such as the demand function and consumer welfare calculations.

Note that the fact that we are assuming a parametric form for the utility function is what allows us to obtain identification of the utility function parameters in a single market. There is no way to obtain higher-order approximations to utility without either observing more choices for each consumer (under different prices) or making additional homogeneity assumptions, in which case the criticisms of Brown and Rosen (1982), Bartik (1987), and Epple (1987) would still apply.[2] Additionally, note that our approach accounts for the endogeneity of prices because we recover the unobserved product characteristics structurally in the first stage, through the independence assumption (and alternatives). In that sense our approach is stylistically similar to the control function approach of Petrin and Train (2004).

While this paper is directed primarily at the hedonics literature, the approach of the paper also has much in common with demand models used in the recent empirical literatures in industrial organization and marketing, beginning with Berry (1994) and Berry et al. (1995) and including Rossi et al. (1996), Nevo (2001), Petrin (2002), and many others. In particular, that literature has emphasized the importance of including unobserved product characteristics in the demand model. The primary difference between the hedonic model in this paper and the Berry et al.–style models is that the hedonic model does not have an independently and identically distributed random error term in the utility function. One consequence of this is that in practice the two models might lead to different results, though we leave a full investigation of that possibility to future research (see also Berry and Pakes [2001], Petrin [2002], and Ackerberg and Rysman [forthcoming] for further discussion of these issues). Berry and Pakes also provide an alternative method to ours for estimating pure hedonic discrete-choice models. Finally, Bajari and Kahn (2005) apply our methods to estimating housing demand and evaluating racial segregation in cities.

In Section VI, we apply our methods to the problem of estimating the demand for personal computers using data that contain 695 products and 19 characteristics. We found that, with these data, the first-stage estimates were surprisingly precise, suggesting that the first-stage

---

[2] Ekeland, Heckman, and Nesheim (2004) provide a solution to the identification problem outlined by Bartik (1987) and Epple (1987). However, their approach allows for only a single-dimensional characteristic that must be observed.

data requirement may not be as high as expected for a nonparametric system. We also investigate the extent to which discreteness of the product space is important for identification of preferences in the second stage. We find that in our data, discreteness of the product space would not be very important in identifying preferences if all the characteristics were continuous (e.g., CPU, RAM, hard drive). On the other hand, we find that the discrete characteristics (digital video disk drive, modem, etc.) are economically important but serve to considerably widen the bounds on both preference parameters and the demand curve. However, while the discrete characteristics do lead to relatively wide bounds on the demand curve, we find that this does not always translate into wide bounds in welfare calculations. For example, when we compute bounds on the compensating variation required if each consumer's first-choice product were removed from the choice set, we find that the average upper bound is $213 and the median upper bound is $59, both numbers that seem reasonable in this context.

In our opinion, the primary disadvantage of the approach presented here is its data requirement. The first-stage estimation is likely to require significantly more data than other alternatives and would place limits on the applicability of the approach in markets with small numbers of products or large numbers of characteristics (without additional assumptions). Another potential disadvantage is that it requires a stronger independence assumption than is commonly used. However, we also relax this requirement somewhat by allowing the error term to be nonseparable. Thus the model is easily capable of generating features such as heteroskedasticity that are often found in empirical work. Finally, the method allows for only a single-dimensional, vertically differentiated unobserved characteristic.[3]

The rest of the paper proceeds as follows. Section II introduces the model and notation and proves that if demand is given by the hedonic model, then there exists an equilibrium price function. Section III shows identification of the price function and unobserved product characteristics (first stage). Section IV shows identification of preferences (second stage). Section V presents econometric estimators consistent with the arguments of Sections III and IV. Finally, Section VI applies the estimators to estimating the demand for personal computers.

## II.    The Model

In our model, a product $j \in \mathcal{J}$ is a finite-dimensional vector of characteristics, $(\boldsymbol{x}_j, \xi_j)$, where $\boldsymbol{x}_j = (x_{j1}, \ldots, x_{jK})$ is a $K$-dimensional vector of

---

[3] Goettler and Shachar (2001) relax both of these assumptions in the Berry et al. framework. Benkard and Bajari (2005) use techniques similar to those of this paper to recover a multidimensional unobserved characteristic in the context of price indexes.

characteristics observed by both the consumer and the econometrician, and $\xi_j$ is a scalar that represents a characteristic of the product that is observed only by the consumer. The unobserved characteristic reflects the fact that there are some characteristics, such as the quality of the good, that are often not observed by the econometrician. We shall assume that each characteristic $x_{jk} \in \mathbb{R}$. However, characteristics may be continuous, such as CPU speed or RAM, or discrete, such as whether or not there is a modem. For our results it is necessary to collapse the unobserved characteristics into a single index.[4] The set $X = \cup_{j \in \mathcal{J}}(\boldsymbol{x}_j, \xi_j) \subseteq R^{K+1}$ represents all products that are available to consumers in the market.

Let $p_{jt}$ denote the price of product $j$ in market $t \in T$. The elements of $T$ can be thought of either as a time series for a single market or as a cross section of markets, or as a panel. Consumers are utility maximizers who select a product $j \in \mathcal{J}$ to consume along with an amount of the composite commodity, $c \in R_+$. Each consumer, $i$, has income $y_i$ to spend each period. The key feature that defines the hedonic model is that consumers' utility functions are defined over characteristics instead of goods. Consumers' utility functions are given by $u_i(\boldsymbol{x}_j, \xi_j, c) : X \times R_+ \rightarrow R$. The price of the composite commodity is normalized to one. Formally, the consumer's maximization problem is

$$\max_{(j,c)} u_i(\boldsymbol{x}_j, \xi_j, c) \quad \text{subject to } p_{jt} + c \leq y_i. \tag{1}$$

Below we shall consider utility functions that have known parametric form

$$u_i(\boldsymbol{x}_j, \xi_j, c) = u(\boldsymbol{x}_j, \xi_j, c; \boldsymbol{\beta}_i).$$

For example, one commonly used functional form is the quasi-linear

$$u(\boldsymbol{x}_j, \xi_j, c; \boldsymbol{\beta}_i) = \log(\boldsymbol{x}_j) \cdot \boldsymbol{\beta}_{i,x} + \log(\xi_j)\beta_{i,\xi} + c,$$

where $\boldsymbol{\beta}_i = (\boldsymbol{\beta}_{i,x}, \beta_{i,\xi})$ represents individual $i$'s vector of taste coefficients. Note that the main difference between the hedonic model and random coefficients versions of standard discrete-choice models is that the hedonic model does not incorporate an error term that is specific to both the choice and the individual (commonly denoted $\epsilon_{ij}$).

## A. The Price Function

In this subsection we investigate the relationship between prices and characteristics in equilibrium if demand is given by the hedonic model

---

[4] Benkard and Bajari (2005) allow for a higher-dimensional unobservable but find that a single index approximates the higher-dimensional case well.

specified above. Rosen (1974) assumes that there is perfect competition and a continuum of products. These assumptions justify the existence of a price function relating characteristics to prices for all possible bundles, also justifying the first step of his estimation procedure in which this price function is estimated. Our goal is to broaden the application of hedonic models to cases in which there may be imperfect competition and a finite number of products, so we shall relax these two assumptions. It turns out that under weak conditions a price function still exists.

Specifically, we now show that in any equilibrium the hedonic model given above implies that prices in each market must have the following properties: (1) there is one price for each bundle of characteristics, (2) the price surface is increasing in the unobserved characteristic, and (3) the price surface satisfies a Lipschitz condition. These properties are important because they make it possible both to estimate the price surface and to use the price surface to recover the unobserved characteristics.

To show that the properties listed above must hold, we rely only on consumer maximization, the fact that prices are taken as given by consumers, and some simple conditions on consumers' utility functions. Most important, because it is the demand side of the model that is restricting prices, we require no supply-side assumptions.

ASSUMPTION 1.

a. $u_i(\boldsymbol{x}_j, \xi_j, c)$ is continuously differentiable in $c$ and strictly increasing in $c$, with $\partial u_i(\boldsymbol{x}_j, \xi_j, c)/\partial c > \epsilon$ for some $\epsilon > 0$ and all $c \in (0, y_i]$.
b. $u_i$ is Lipschitz continuous in $(\boldsymbol{x}_j, \xi_j)$.
c. $u_i$ is strictly increasing in $\xi_j$.

The first two conditions are quite weak. Condition $c$ implies that there is no satiation in the unobserved product characteristic. This is not guaranteed to hold in all applications, but we do feel that the assumption is usually justifiable in practice. In practice, the unobserved characteristic in the model represents a composite of all unobserved features of the product that we typically associate with hard-to-measure features such as "quality," "style," or "service." These are characteristics that we can comfortably say that the vast majority of individuals would always prefer more of (if more were free). In cases in which condition $c$ does not hold, the model would instead be an approximation to the true model.[5]

THEOREM 1.    Suppose that assumption 1 holds for every individual

[5] Note also that in any model incorporating unobserved characteristics, some assumption such as condition $c$ of assumption 1 is required regardless of the approach used. For example, assumption 1 is equivalent to the monotonicity assumption of Berry (1994).

in a market, $t$. Then, for any two products $j$ and $j'$ with positive demand in market $t$,

1. if $x_j = x_{j'}$ and $\xi_j = \xi_{j'}$, then $p_{jt} = p_{j't}$;
2. if $x_j = x_{j'}$ and $\xi_j > \xi_{j'}$, then $p_{jt} > p_{j't}$;
3. $|p_{jt} - p_{j't}| \leq M(|x_j - x_{j'}| + |\xi_j - \xi_{j'}|)$ for some $M < \infty$.

*Proof.* See the Appendix.

We denote the equilibrium price function for market $t$ as $\mathbf{p}_t(x_j, \xi_j)$. 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 properties 1–3 of theorem 1 hold. Because property 3 holds for all pairs of products, in the limit the price function must be Lipschitz continuous.

Note that the price function in each market is an equilibrium function that is dependent on market primitives. It does not tell what the price of a good would be if that good were 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 or marginal production costs, or if a good already in the market were to be produced by another multiproduct 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 $t$ subscript). 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. This information is helpful in estimating the consumer's preferences.

We write the price function nonseparably because we have very little information about its form a priori except properties 1–3. Its form depends on the primitives of the model such as preferences, marginal costs, and the nature of competition. For example, even a standard single-product firm's inverse elasticity markup formula with a linear marginal cost function would often imply a nonseparable price function. Thus we feel that it would not be appropriate to proceed by making an arbitrary assumption such as additive separability even though such an assumption would make the estimation substantially simpler.

The importance of the theorem can be seen by the fact that there need not be an equilibrium price surface in all demand-supply systems. For example, if demand were given by a logit model (or any standard discrete-choice econometric model), then it would be possible for two different products with the same characteristics to have different prices. This could happen if two firms produced products with identical characteristics but the firms had different costs of production. It could also happen if two multiproduct firms owned different sets of complemen-

tary products. In standard discrete-choice econometric models of demand, two products with identical characteristics but different prices are both guaranteed to have positive demand, so the demand side places less discipline on prices. Properties 1–3 are guaranteed to hold only in the hedonic model.

Note that, because the theorem is based on demand-side arguments only, it is general to many types of equilibria. An advantage of this is that our estimates of preferences below are robust to a variety of supply-side assumptions, including even models of dynamic pricing. Of course, there are also disadvantages to not fully specifying the supply side of the model. A correctly specified supply model would add efficiency to the estimation. Additionally, a model of supply would be required for evaluating counterfactual policies that involve new pricing equilibria.

### III.  Identification of the Price Function and the Unobserved Characteristics

In this section we consider three potential approaches for obtaining identification of the price function and unobserved characteristics. We expect the first method, which uses independence, to be the most useful in practice. The independence assumption we use here is similar to, but slightly stronger than, the independence assumption that is widely used in the empirical literatures in industrial organization and marketing. The second method uses a nonseparable analogue to fixed effects. The third method is useful when independence is questioned, but there are instruments available for the endogenous characteristics. We shall assume throughout that prices are observed without error. However, with some additional assumptions, identification can still be obtained if prices are observed with error.

In a slight abuse of notation, throughout this section we shall denote the joint distribution of the vector of characteristics and prices in market $t$, $(p_{jt}, x_j, \xi_j)$, as $F(p_t, x, \xi)$, with density $f(p_t, x, \xi)$.

### A.  *Identification Using Independence*

The main assumption of this section is that the observed characteristics $x$ are independent of the unobserved characteristic, $\xi$.

ASSUMPTION 2.    $\xi$ is independent of $x$.

Assumption 2 is a strengthening of the mean independence assumption commonly used in the empirical literature in industrial organization. However, allowing for nonseparability makes the implications of the independence assumption less restrictive. In models in which the errors enter linearly, independence rules out heteroskedastic unobservables, which are often thought to be important in applied work.

Nonseparability allows the underlying independent error to interact with the observed data in ways that replicate models of heteroskedasticity. 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 $\xi$.

Since the unobserved product attribute has no inherent units, it is possible to identify it only up to a monotonic transformation. Thus, without loss of generality, we assume that a normalization has been made to $\xi$ such that the marginal distribution of $\xi$ is $U[0, 1]$. Technically, this amounts to normalizing $\xi$ using its distribution function.[6]

Our identification proof, which is constructive, is similar to that in Matzkin (2003). We define identification as in that paper, relative to the normalization above.

THEOREM 2.    Suppose that assumptions 1 and 2 hold. If many products are observed in a single market, then $\{\xi_j\}$ is identified. In each market $t$ in which many products are observed, the price function $\mathbf{p}_t$ is identified.

*Proof.*    We first show how to construct the unobserved product characteristics using the conditional distribution of prices in a single market $t$. The proof assumes that many products are observed in market $t$ such that $F(p_t, x, \xi)$ is known:

$$
\begin{aligned}
F_{p_t|x=x_j}(p_{jt}) &= \Pr\left[\mathbf{p}_t(x, \xi) \leq p_{jt}|x = x_j\right] \\
&= \Pr\left[\xi \leq \mathbf{p}_t^{-1}(x, \, p_{jt})|x = x_j\right] \\
&= \Pr\left[\xi \leq \mathbf{p}_t^{-1}(x_j, \, p_{jt})\right] \\
&= \mathbf{p}_t^{-1}(x_j, \, p_{jt}) \\
&= \xi_j.
\end{aligned}
$$

The second line of the proof holds by property 3 of theorem 1, the third line holds by independence, and the fourth line holds because of the $U[0, 1]$ normalization.

To construct the price function for market $t$, we need only invert the above relationship:

$$
\mathbf{p}_t(x_0, \, e_0) = F_{p_t|x=x_0}^{-1}(e_0). \tag{2}
$$

---

[6] We thank an anonymous referee for pointing out that a drawback of this normalization is that it requires that $\xi$ have a continuous distribution, an additional restriction over the assumptions in the paper. However, this restriction, while convenient, is not necessary for the identification proof. An alternative normalization that does not require that $\xi$ have a continuous distribution is to normalize $\xi$ such that $F(\xi) = F(p|x = \bar{x})$ for some value $\bar{x} \in A$ (see also Matzkin 2003).

This is possible for all markets $t$ in which many products are observed such that $F(p_t, \boldsymbol{x}, \xi)$ is known. QED

Since identification of the unobserved product characteristics is possible from data on a single market, identification is obtained even if the unobserved product characteristics for each product change over time, as is commonly assumed in the empirical literature.

### B.    *Identification Using "Options Packages"*

Because the characteristics of products are jointly chosen by the firm, we may worry that the independence assumption above is violated in some applications. One alternative that avoids this assumption is to use "options packages." 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, etc.) and a package of options associated with the model (horsepower, air conditioning, power steering, etc.). 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, etc.) and an options package (RAM, processor speed/type, hard drive, etc.). Suppose that the unobserved characteristics ($\xi_j$) correspond to a model and the observed characteristics ($\boldsymbol{x}_j$) correspond to an options package. In that case, we show that the price data across options packages for each model can be used to identify the unobserved characteristics of the different models.

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 $\pi$ is the set of products that are model $z$, where each product has a possibly different options package $\boldsymbol{x}$. The model $z$ is observable, and $\boldsymbol{x}$ and $z$ have joint distribution $F_{x,z} : A \times \mathcal{Z} \to R$.

The first assumption in this subsection says that $\xi$ is shared by products that are the same model.

ASSUMPTION 3.    For all $j_1, j_2 \in \mathcal{J}$, 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.

ASSUMPTION 4.    There exists an $\bar{\boldsymbol{x}} \in A$ such that, for all $z \in \mathcal{Z}$, $f(\bar{\boldsymbol{x}}|z) > 0$.

Because of the lack of implicit units for $\xi$, we again can identify $\xi$ and the price function only up to a normalization. In this case we normalize $\xi$ such that $F_{\xi|x=\bar{x}}$ is $U[0, 1]$.

THEOREM 2.    Suppose that assumptions 1, 3, and 4 hold. If many products are observed in a single market, then $\{\xi_j\}$ is identified. In each

market $t$ in which many products are observed, the price function $\mathbf{p}_t$ is identified.

*Proof.* For each product $j$, let $j^*$ be a product such that $\pi(j) = \pi(j^*)$ and $\mathbf{x}_{j^*} = \bar{\mathbf{x}}$. Such a product exists for every model $\pi(j)$ by assumption 4. Then, similarly to the previous section,

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

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

The price function in each market $t$ is given by the prices of non-baseline packages. For any point $(\mathbf{x}_0, e_0) \in A \times E$,

$$\mathbf{p}_t(\mathbf{x}_0, e_0) = p_{kt} \quad \text{for } k \in \mathcal{J} \tag{3}$$

such that $\xi_k = e_0$ and $\mathbf{x}_k = \mathbf{x}_0$. QED

Again in this case, identification of the unobserved product characteristics is obtained in a single cross section. Note also that since many products are observed in each market that have the same value for the unobservable, the model is overidentified and is therefore testable.

## C.   *Identification Using Instruments*

In the event that the unobserved product characteristic is not independent of all the observed characteristics, a second alternative would be to use nonseparable instrumental variables techniques (e.g., Chesher [2003], Imbens and Newey [2003], Chernozhukov, Imbens, and Newey [2004], and Chernozhukov and Hansen [2005]; our approach most closely relates to that in Imbens and Newey). For example, Bayer, McMillan, and Rueben (2004) use the fixed characteristics (e.g., housing stock characteristics) of housing in surrounding neighborhoods as instruments for the endogenous characteristics (e.g., racial makeup) of a given neighborhood.

Suppose that one of the observed characteristics is endogenous in the sense that it may be correlated with (or otherwise depends on) the unobserved characteristic. Without loss of generality, we assume that this is the first characteristic, $x_1$. We also suppose that some instruments, $\mathbf{z}$, are available such that[7]

$$x_1 = h(\mathbf{x}_{-1}, \mathbf{z}, \eta).$$

This instrumenting equation is analogous to a reduced form for the characteristic $x_1$. Similarly to linear instrumental variables approaches, the instruments must be jointly independent of the unobserved characteristics, $\xi$, and the error term in the instrumenting equation, $\eta$.

---

[7] We allow $x_1$ to possibly depend on the other observed characteristics, but this is not required for identification.

ASSUMPTION 5.    $(\xi, \eta)$ are jointly independent of $(\mathbf{x}_{-1}, \mathbf{z})$.

Thus the endogeneity of $x_1$ results from a correlation between $\xi$ and $\eta$. We also require monotonicity of the instrumenting equation in the error term.

ASSUMPTION 6.    $h(\mathbf{x}_{-1}, \mathbf{z}, \cdot)$ is strictly increasing.

Assumptions 5 and 6 make the instrumenting equation analogous to the price equation in subsection $A$. By the results of that subsection, the function $h(\cdot)$ and the error $\eta$ are identified. Furthermore, if $\eta$ is normalized by its distribution function such that it is $U[0, 1]$, then we can retrieve $\eta$ using the relationship $\eta = F(x_1|\mathbf{x}_{-1}, \mathbf{z})$.

Once $\eta$ is known, it is straightforward to show that the price function is identified. To do this we rely on an observation by Imbens and Newey (2003) that $x_1$ and $\xi$ are independent conditional on $\eta$. This can be seen as follows. Since $(\eta, \xi)$ are jointly independent of $(\mathbf{x}_{-1}, \mathbf{z})$, $\xi$ is independent of $(\mathbf{x}_{-1}, \mathbf{z})$ conditional on $\eta$. Thus $\xi$ is independent of $h(\mathbf{x}_{-1}, \mathbf{z}, \eta) = x_1$ conditional on $\eta$.

THEOREM 4.    Suppose that assumptions 1, 5, and 6 hold. If many products are observed in a single market $t$, then $\{\xi_j\}$ is identified. In each market $t$ in which many products are observed, the price function $\mathbf{p}_t$ is identified.

*Proof.*    The error term $\eta$ and $h(\cdot)$ are identified from data on a single market by theorem 2. For a single market $t$,

$$\Pr\left(p \leq p_{jt}|\mathbf{x} = \mathbf{x}_j, \eta\right) = \Pr\left[\mathbf{p}_t(\mathbf{x}, \xi) \leq p_{jt}|\mathbf{x} = \mathbf{x}_j, \eta\right]$$
$$= \Pr\left[\xi \leq \mathbf{p}_t^{-1}(\mathbf{x}, p_{jt})|\mathbf{x} = \mathbf{x}_j, \eta\right]$$
$$= \Pr\left[\xi \leq \mathbf{p}_t^{-1}(\mathbf{x}_j, p_{jt})|\eta\right],$$

where the last equality follows by the independence result in the text above. Integrating both sides with respect to $\eta$ (which has unit density given the normalization) gives

$$\int \Pr\left(p \leq p_{jt}|\mathbf{x} = \mathbf{x}_j, \eta\right)d\eta = \Pr\left[\xi \leq \mathbf{p}_t^{-1}(\mathbf{x}_j, p_{jt})\right]$$
$$= \mathbf{p}_t^{-1}(\mathbf{x}_j, p_{jt})$$
$$= \xi_j. \tag{4}$$

Since $\eta$ is identified, the expression on the left-hand side is known. This shows identification of $\xi_j$. The function $\mathbf{p}_t$ is obtained for any market $t$ by inverting the left-hand-side expression. QED

Note that again the unobserved product characteristics are identified in a single cross section.

### IV. Identification of Preferences

In this section we consider the identification of consumer preferences, the demand function, and consumer surplus. We begin by outlining the simplest case: the choice set is continuous. This case is similar to that of Rosen (1974) and would apply best to markets such as housing that have very rich product spaces. Many applications in the recent literature in industrial organization have involved markets in which choices are more limited, such as autos and packaged goods. Therefore, we also consider identification in that case below. Throughout, we assume that only a small number of observations are available per individual since this is most often the case in applications.

#### A. *Continuous Choice Set*

When the choice set is continuous, by standard arguments, at each bundle chosen by a consumer, the marginal rate of substitution between characteristics must be equal to the slope of the consumer's budget set at that point. Since the consumer's budget set is known completely from the function $\mathbf{p}_t(x, \xi)$, observation of a consumer's chosen bundle is sufficient to learn the consumer's marginal rate of substitution at that bundle. However, this information provides only local information about preferences at each chosen bundle; so if only a small number of choices are observed per consumer, identification of the consumer's weak preference relation is not possible. Instead, following recent convention in the empirical literatures in industrial organization and marketing, we proceed using parametric restrictions to the consumer's indifference curves.

In the empirical literature it is common to specify the utility function as being either linear or log-linear in the characteristics, for example,

$$u_{ij} = \beta_{i,1} \log (x_{j,1}) + \cdots + \beta_{i,K} \log (x_{j,K}) + \beta_{i,\xi} \log (\xi_j) + c. \tag{5}$$

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

$$\frac{\beta_{i,k}}{x_{j,k}} = \frac{\partial \mathbf{p}_t}{\partial x_{j,k}} \quad \text{for } k = 1, \ldots, K \tag{6}$$

and

$$\frac{\beta_{i,\xi}}{\xi_j} = \frac{\partial \mathbf{p}_t}{\partial \xi_j}. \tag{7}$$

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

$$\beta_{i,k} = x_{j,k}\frac{\partial \mathbf{p}_t}{\partial x_{j,k}} \quad \text{for } k = 1, \ldots, K \tag{8}$$

and

$$\beta_{i,\xi} = \xi_j\frac{\partial \mathbf{p}_t}{\partial \xi_j}. \tag{9}$$

If the price function, $\mathbf{p}_t$, and unobserved characteristics $\{\xi_j\}$ are known, then in this example household $i$'s preference parameters, $\boldsymbol{\beta}_i = (\beta_{i,1}, \ldots, \beta_{i,K}, \beta_{i,\xi})$, can be recovered even if only a single choice of the household, $(\mathbf{x}_j, \xi_j)$, is observed. By aggregating the decisions of all the households in market $t$, one can learn the population distribution of taste coefficients in market $t$, $G_t(\boldsymbol{\beta})$.

Of course, as more choices are observed, it is possible to estimate more complex functional forms for utility. In general, we characterize an agent by a $B$-dimensional parameter vector $\boldsymbol{\beta}_i \in R^B$. Since the previous section has shown that the unobservables, $\{\xi_j\}$, are identified by the price function, we proceed as though $\xi$ is known and write the utility function as

$$u_i(\mathbf{x}, c) = u(\mathbf{x}, y_i - \mathbf{p}(\mathbf{x}); \boldsymbol{\beta}_i), \tag{10}$$

where the dependence of utility on $\xi$ is dropped to simplify notation.

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

$$\frac{\partial}{\partial x_k}[u(\mathbf{x}, y_i - \mathbf{p}(\mathbf{x}); \boldsymbol{\beta}_i)] = 0 \quad \text{for } k = 1, \ldots, K. \tag{11}$$

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

$$\mathbf{x}'(\boldsymbol{\beta}) = -(D_{\mathbf{x},\mathbf{x}}\bar{u})^{-1}D_{\mathbf{x},\boldsymbol{\beta}_i}\bar{u}, \tag{12}$$

where

$$\bar{u}(\mathbf{x}; \boldsymbol{\beta}) = u(\mathbf{x}, y_i - \mathbf{p}(\mathbf{x}); \boldsymbol{\beta}_i). \tag{13}$$

THEOREM 5.    Suppose that $\boldsymbol{\beta}_i \in \mathcal{B} \subseteq R^B$, where $\mathcal{B}$ is an open convex subset and $\mathbf{x} \in R^K$. If $\mathbf{x}'(\boldsymbol{\beta})$ is locally negative definite or positive definite, then $\boldsymbol{\beta}_i$ is locally identified. If $K = B$ and $\mathbf{x}'(\boldsymbol{\beta})$ is globally positive definite or negative definite, then $\mathbf{x}(\boldsymbol{\beta})$ 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 (12) is everywhere positive or negative definite, then $x(\boldsymbol{\beta})$ is one-to-one so that the preferences are globally identified (see Gale and Nikaidô 1965). QED

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, **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. If the choice set is rich, many welfare experiments (e.g., evaluating the effect of a price change) would involve only local changes to utility. 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.

### B.    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 (11), when the product space is discrete, consumer maximization implies a set of inequality constraints. If consumer $i$ chooses product $j \in \{1, \dots, J\}$, then

$$u(x_j, \xi_j, y_i - p_j;\ \boldsymbol{\beta}_i) \geq u(x_k, \xi_k, y_i - p_j;\ \boldsymbol{\beta}_i) \quad \text{for all } k \neq j. \quad (14)$$

Therefore, it must be that $\boldsymbol{\beta}_i \in A_i$, where

$$A_i = \{\boldsymbol{\beta}_i : \boldsymbol{\beta}_i \text{ satisfies } (14)\}. \quad (15)$$

If the choice set is finite, the $A_i$ sets will typically not be singletons, implying that the parameters $\boldsymbol{\beta}_i$ are not point-identified. However, that does not mean that the data are noninformative. If the choice set is rich, the $A_i$ sets may be small. Furthermore, it is easy to show that if all the characteristics are continuous and the choice set is compact, then as the number of products increases, the $A_i$ sets converge to the individual taste coefficients $\boldsymbol{\beta}_i$.

In applications in which the $A_i$ sets are large enough that the lack of point identification matters practically, we suggest two alternative approaches. The first is to proceed in the spirit of the recent literature on set-identified models (Manski and Tamer 2002; Haile and Tamer 2003) and construct bounds on the preference distributions. The $A_i$

sets represent bounds on the preference parameters for each consumer. Together, they can be used to construct bounds on the distribution of individual preference parameters in the population.

Suppose that there are $M$ consumers in the market. Then the aggregate distribution of preference parameters is defined as

$$G_t(\boldsymbol{\beta}) = \frac{1}{M}\sum_{i=1}^{M} \mathbf{1}\{\boldsymbol{\beta}_i \leq \boldsymbol{\beta}\},$$

where $\mathbf{1}\{\boldsymbol{\beta}_i \leq \boldsymbol{\beta}\}$ denotes an indicator function that is one if individual $i$'s preference parameter is less than (in the vector sense) the vector $\boldsymbol{\beta}$. Bounds on $G_t(\boldsymbol{\beta})$ can be constructed using the $A_i$ sets as follows:[8]

$$G_t(\boldsymbol{\beta}) \leq \overline{G}_t(\boldsymbol{\beta}) \equiv \frac{1}{M}\sum_{i=1}^{M}\sup_{\boldsymbol{b}\in A_i} \mathbf{1}\{\boldsymbol{b} \leq \boldsymbol{\beta}\} \tag{16}$$

and

$$G_t(\boldsymbol{\beta}) \geq \underline{G}_t(\boldsymbol{\beta}) \equiv \frac{1}{M}\sum_{i=1}^{M}\inf_{\boldsymbol{b}\in A_i} \mathbf{1}\{\boldsymbol{b} \leq \boldsymbol{\beta}\}. \tag{17}$$

The intuition of the upper bound is that if there is any vector $\boldsymbol{b}$ in the set $A_i$ such that $\boldsymbol{b} \leq \boldsymbol{\beta}$, then it is possible that consumer $i$'s preference parameter is less than $\boldsymbol{\beta}$. The upper bound is chosen so as to encompass all such possibilities (with the lower bound chosen similarly).

An alternative way to address the lack of point identification would be to use a parametric approximation to the preference distribution. While each individual's preference parameters are not point-identified from his or her choices, putting together the information contained in all individuals' choices should still provide a lot of information about the aggregate preference distribution. Thus, depending on how flexibly the aggregate preference distribution is modeled and how rich the data are, the parameters of the preference distribution as a whole are still likely to be point-identified (see Bajari and Kahn [2005] for an example).

### C.    Nonpurchasers and Outside Goods

In our model, individuals who 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 her income on the composite commodity $c$. That is, it is as though 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

---

[8] Note the slightly confusing feature of the bounds that a lower (upper) bound for the distribution function corresponds to an upper (lower) bound for the distribution itself.

would imply a set of inequalities for nonpurchasers of the form

$$u(0, 0, y_i; \boldsymbol{\beta}_i) \geq u(\boldsymbol{x}_k, \xi_k, y_i - p_k; \boldsymbol{\beta}_i) \quad \text{for all } k. \tag{18}$$

These inequalities could then be used similarly to those above in (14) to locate nonpurchasers' preference parameters. Note that (18) provides only inequalities and therefore provides only set identification for nonpurchasers' preference parameters even if the choice set is continuous.

### D. Demand and Welfare

Once consumers' preference parameters have been recovered, it is straightforward to compute the demand system. Demand for product $j$ as a function of the prices of all the goods is given by

$$q_j(p_1, \ldots, p_J) = \sum_{i=1}^{M} \mathbf{1}\{u_{ij} \geq u_{ik} \text{ for all } k \neq j\}.$$

In the event that it is possible to recover only bounds on individual consumers' preference parameters, demand cannot be computed exactly. However, the preference parameter bounds can be used to compute bounds on demand:

$$q_j(p_1, \ldots, p_J) \leq \bar{q}_j \equiv \sum_{i=1}^{M} \sup_{\boldsymbol{b} \in A_i} \mathbf{1}\{u(\boldsymbol{x}_j, \xi_j, y_i - p_j; \boldsymbol{b})$$

$$\geq u(\boldsymbol{x}_k, \xi_k, y_i - p_k; \boldsymbol{b}) \text{ for all } k \neq j\} \tag{19}$$

and

$$q_j(p_1, \ldots, p_J) \leq \underline{q}_j \equiv \sum_{i=1}^{M} \inf_{\boldsymbol{b} \in A_i} \mathbf{1}\{u(\boldsymbol{x}_j, \xi_j, y_i - p_j; \boldsymbol{b})$$

$$\geq u(\boldsymbol{x}_k, \xi_k, y_i - p_k; \boldsymbol{b}) \text{ for all } k \neq j\}. \tag{20}$$

Note that demand is determined by overall willingness to pay for a product and not the particular values of any one preference parameter. Thus it is possible for wide bounds on the preference parameters to imply narrow bounds on demand (and also vice versa).

It is also straightforward to do welfare calculations and to compute bounds on consumer welfare. For example, suppose that we want to compute the welfare effects of a change in prices or the choice set for a consumer with income $y$ and preference parameters $\boldsymbol{\beta}$. Let $\mathcal{C}_1 \equiv \{(\boldsymbol{x}_j, \xi_j, p_j)\}_{j=1}^{J_2}$ denote the consumer's initial choice set and the initial prices and $\mathcal{C}_2 \equiv \{(\boldsymbol{x}_j, \xi_j, p_j)\}_{j=1}^{J_2}$ denote the new choice set and prices.

Utility received by the consumer from the initial choice set is

$$u_1^* = \max_{j \in C_1} u(\boldsymbol{x}_j, \xi_j, y - p_j; \boldsymbol{\beta}).$$

In order to make the consumer equally well off when facing the choice set $C_2$, the consumer needs to be compensated by an amount $W$ such that

$$\max_{j \in C_2} u(\boldsymbol{x}_j, \xi_j, y - p_j + W; \boldsymbol{\beta}) = u_1^*.$$

If the consumer's preference parameters are only set-identified, then we can compute bounds on this compensating variation, $W$, by taking the sup and the inf of $W$ as calculated above for $\boldsymbol{\beta} \in A_i$.

## V.  Estimation

In all the cases considered above the identified objects have empirical counterparts that are straightforward to calculate and use for estimation. We assume that the econometrician observes prices and characteristics for $j = 1, \ldots, J$ products across $t = 1, \ldots, T$ markets. We also assume that the choices of a sample of $N$ consumers are observed. We allow that the choice data may be aggregate data in the sense that no information is available matching consumers' demographics to purchases. In that case the choice data would include only the total market demand for each product.

### A.  First-Stage Estimation

If the choice set is discrete, the first stage consists only of estimating the value of the unobservables. If the choice set is continuous, it is also necessary to know the price function derivatives. If there is measurement error in prices, then before the first-stage estimation it is necessary to do some smoothing to remove the measurement error.[9]

### 1.  Independence Case

In this subsection we assume that $\boldsymbol{x}$ and $\xi$ are independent. As in equation (2), in this case all that is required for estimation of both $\xi$ and the price function is an estimator for the conditional distribution of prices given characteristics, $\hat{F}_{p_t|\boldsymbol{x}=\boldsymbol{x}_0}(e_0)$. For example, a kernel estimator (such as those outlined in Matzkin [2003]) or a series estimator

---

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

(such as those outlined in Imbens and Newey [2003]) could be used. In applying these estimators to our data on personal computers, we found that a local linear kernel estimator (Fan and Gijbels 1996) worked best.

The unobserved characteristics are obtained from the estimator directly,

$$\hat{\xi}_{jt} = \hat{F}_{p_t|x=x_j}(p_{jt}),\tag{21}$$

whereas the price function is obtained by inverting the estimator.

While Matzkin (2003) does not explicitly consider estimation of the unobservable, the asymptotic properties of the estimator in (21) are analogous to those of the estimator considered in theorem 4 of that paper.

## 2.  Options Packages Case

As in the independence case, all that is required here is an estimator of the distribution of prices conditional on characteristics, $\hat{F}_{p_t|x=x_0}(e_0)$. However, in the options packages case, this estimator is evaluated only at $x = \bar{x}$ when estimating $\xi$,

$$\hat{\xi}_{jt} = \hat{F}_{p_t|x=\bar{x}}(p_{j*t}),\tag{22}$$

where $j*$ is the product that is the same model as $j$ with options package $x = \bar{x}$. If such a package is not actually observed for all products, it is first necessary to "predict" $p_{j*t}$ by smoothing the prices of products that are the same model as $j$ and that have characteristics close to $\bar{x}$.

## 3.  Instrumental Variables Case

In the instrumental variables case, it is first necessary to estimate the instrumenting equation. The error in the instrumenting equation, $\eta$, is estimated using the distribution of $x_1$ conditional on $x_{-1}$ and $z$:

$$\hat{\eta}_{jt} = \hat{F}_{x_1|x_{-1},z}(x_{1j}).$$

This estimator is analogous to the estimation of $\xi$ in the independence case above.

Next, it is necessary to estimate the distribution of prices conditional on $x$ and $\eta$:

$$\hat{F}_{p_t|x=x_j,\eta=\eta_{jt}}.\tag{23}$$

This conditional distribution then has to be integrated over $\eta$ as in (4) to obtain an estimator of $\xi$:

$$\hat{\xi}_{jt} = \int \hat{F}_{p_t | x = x_j, \eta}(p_{jt}) d\eta.$$

## B.  *Estimation of Preferences, Continuous Product Space*

Here we consider estimating the preference parameters when there is only one choice observation per consumer 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, suppose that the consumer's utility takes the form in equation (5). Then the first-order conditions imply that equations (8) and (9) must hold. This suggests the following estimator for $\beta_i$:

$$\hat{\beta}_{i,k} = x_{j,k} \frac{\partial \hat{\mathbf{p}}_t}{\partial x_{j,k}} \quad \text{for } k = 1, \ldots, K \tag{24}$$

and

$$\hat{\beta}_{i,\xi} = \hat{\xi}_j \frac{\partial \hat{\mathbf{p}}_t}{\partial \xi_j}, \tag{25}$$

where $(x_j^i, \hat{\xi}_j^i)$ represents the (estimated) bundle chosen by individual $i$ and $\partial \hat{\mathbf{p}}_t / \partial x_{j,k}$ represents an estimator for the derivative of the price function at the chosen bundle. One way to estimate the price function derivatives is to use the derivatives of a price function estimator. The price function can be estimated analogously to (21) above (except using [2]) and using either a kernel or series-based approach. Matzkin (2003) also provides a direct estimator for the derivatives of a nonseparable function.

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, one is then able to construct a density estimate of the joint distribution of taste coefficients and demographics in the population.

*C.    Estimation of Preferences, Discrete Product Space*

Estimation is even more straightforward if the product space is discrete. The $A_i$ sets can be estimated for each consumer by plugging the consumer's observed choice and the estimated values of $\xi$ for each product into the inequalities (14) and then solving for the set of $\beta$'s that satisfy them. Operationally this can be done in several ways. The easiest way, and the method we use in the application in the next section, is to use simulation.[10] Moreover, since the functional forms that we consider are linear in the parameters, the set of inequalities is a linear system, which is a very well studied problem. Motzkin et al. (1953) provide an efficient solution method for such systems, and code for implementing their algorithm is widely available on the Internet. Pakes et al. (2005) also provide a method for solving such systems using linear programming. Another appealing feature of linear systems is that the $A_i$ sets are well behaved (they are convex polyhedrons).

Alternatively, it may often be desirable to estimate the distribution of preferences parametrically. In that case, for each value of the parameter we need to compute the probabilities attached to the $A_i$ sets under the distribution function. These probabilities can be used to construct a likelihood function or market shares for a method of moments estimator. Again, simulation would typically be the easiest way to compute these probabilities. However, we note that this approach often yields an objective function that is not well behaved and that may be difficult to maximize except in smaller-dimensional problems. To solve this problem, in the NBER working paper version of this paper (Benkard and Bajari 2004, 21–25), we provided a computationally simple Gibbs sampling estimator that provides a direct estimate of the aggregate preference distribution. We refer readers to that paper for more details.

## VI.    Application to Computer Demand

In this section we apply our proposed estimators to the demand for desktop personal computers. Our data come from the *PC Data Retail Hardware Monthly Report* and include quantity sold, average sales price, and a long list of machine characteristics for desktop computers. Please see Benkard and Bajari (2005) for a more detailed discussion of these data.

The raw data set contained 29 months of data, but we initially concentrate only on the data for the last period, December 1999, covering 695 machines. We chose to use data for a single period both to keep the estimation simple and also to test how well our methods work for

---

[10] Please see the NBER working paper version of this paper (Benkard and Bajari 2004, 21–25) for details on how this simulation is performed.

TABLE 1
SUMMARY STATISTICS

| Variable | Mean | Standard Deviation | Minimum | Maximum | OLS Coefficient |
|---|---|---|---|---|---|
| CPU benchmark | 1,354.5 | 362.3 | 516 | 2,544 | .836 |
| RAM (megabytes) | 74.0 | 35.1 | 16 | 256 | 3.010 |
| Hard drive (megabytes) | 9,276.8 | 4,850.3 | 2,100 | 40,000 | .008 |
| MMX | .64 | .48 | 0 | 1 | −56.971 |
| SCSI | .01 | .08 | 0 | 1 | 310.747 |
| CD-ROM | .67 | .47 | 0 | 1 | 26.478 |
| DVD | .14 | .35 | 0 | 1 | 32.213 |
| NIC | .36 | .48 | 0 | 1 | 9.481 |
| Monitor? | .31 | .46 | 0 | 1 | 29.625 |
| Monitor size | .75 | 3.27 | 0 | 15 | 22.822 |
| ZIP | .05 | .22 | 0 | 1 | 20.440 |
| Desktop | .17 | .37 | 0 | 1 | 25.611 |
| Refurbished | .09 | .28 | 0 | 1 | −144.314 |
| No modem | .55 | .50 | 0 | 1 | 145.169 |
| Win NT 4.0 | .02 | .14 | 0 | 1 | −106.374 |
| Win NT | .17 | .37 | 0 | 1 | 22.567 |
| Win 98 | .58 | .49 | 0 | 1 | −59.590 |
| Win 95* | .16 | .37 | 0 | 1 | −42.058 |
| Constant | | | | | −590.2 |
| $R^2$ | | | | | .79 |
| Observations | | | | | 695 |

* Win 3.1 is omitted.

this case. The data set reportedly covers approximately 75 percent 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 variable.[11] The final data set contained 19 characteristics, including five operating system dummies (Win 3.1, NT 4.0, NT, Win 98, and Win 95) plus CPU benchmark, multimedia extensions, RAM, hard drive capacity, small-computer system interface, CD-ROM, DVD, modem, modem speed, network interface card, monitor dummy, monitor size (if monitor supplied), ZIP drive, desktop (vs. tower), and refurbished. Summary statistics for these data are given in table 1. To show that accounting for unobserved characteristics is very important in this application, under the assumption that all characteristics are perfectly observed, 357 of the 695 products are dominated, meaning that there is some other product that has strictly better values of the observed characteristics and a lower price. Hence, a hedonic model that did not allow for unobserved characteristics would not be able to rationalize positive demand for more than half of the products observed in the data.

[11] The CPU benchmark was obtained from http://www.cpuscorecard.com. A regression of CPU benchmark on chip speed interacted with chip dummies yielded an $R^2$ of 0.999.

*A. 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. Thus the first-stage estimates are obtained using price data after the linear effect of the other 16 characteristics has been removed (ordinary least squares [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 bought and installed separately. Arbitrage between the personal computer and peripheral markets should place a limit on 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 (21), with a normal kernel. Equation (21) was derived using the $U[0, 1]$ normalization and thus results in estimates of $\xi$ 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 distribution function somewhat relative to the mean square error minimizing bandwidth. Note, however, that because we are treating the product space as finite, we need only the first-stage estimates of $\xi$ in order to estimate the demand system. We do not use the estimates of the price function directly.

An important issue with respect to our approach is how precisely it is possible to estimate the $\xi$'s for each product. Table 2 shows the distribution of standard errors for the estimated $\xi$'s using both an asymptotic approximation (from Fan and Gijbels [1996]) and also a bootstrap method. The asymptotic standard errors average 0.01, with 95 percent of the estimated $\xi$'s having standard errors less than 0.018. The bootstrap standard errors were generally very similar but slightly larger in magnitude. The table also includes a third column that adds in the estimation error from the 16 OLS coefficients. Because some of these coefficients are not estimated well (because of the small numbers of machines with certain characteristics), these standard errors are somewhat larger, averaging 0.08.

In summary, we found that, with very few exceptions, it was possible

TABLE 2
Distribution of Standard Errors for Estimates of $\xi$ ($N=695$)

| Quantile | Asymptotic (1) | Bootstrap (1,000 Samples) (2) | Bootstrap (OLS Coefficients) (1,000 Samples) (3) |
|---|---|---|---|
| Minimum | .002 | .002 | .004 |
| .30 | .007 | .007 | .073 |
| .50 | .009 | .009 | .081 |
| .95 | .018 | .022 | .105 |
| .99 | .052 | .091 | .119 |
| Maximum | .104 | .121 | .153 |
| Average | .010 | .012 | .080 |

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 $\xi$ 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 $\xi$ estimates and found it to be small; but more important, we found that the rank ordering of the $\xi$ estimates was nearly identical whether or not the estimated bias term was included. Hence, we left the estimated bias out of our final first-stage estimates.

### B.  Second-Stage Results: Preferences and Demand Curves

In estimation of the preference parameters and the demand system, high-dimensional systems are easily handled. Therefore, we evaluate two different versions of the demand model. The first (model 1) uses only the continuous characteristics (CPU, hard drive, RAM, and $\xi$) as in the first-stage estimation above. Using this model, we can evaluate the extent to which the discreteness of the product space affects identification of preferences and demand even when all the characteristics are continuous. In the second model (model 2) we use all 19 characteristics, including all the discrete characteristics. In both models we use the quasi-linear functional form for utility (as in eq. [5]), linear in the logs of the continuous characteristics and linear in the remaining characteristics, including expenditure on other goods (price). The unobserved characteristics obtained from the first stage were normalized such that budget sets were always convex. We also tried several alternative nor-

TABLE 3
Characteristics of Four Representative Products Available in December 1999

| Product Number | Brand/Model | CPU (Benchmark) | RAM | Hard Drive | Price | Sales |
|---|---|---|---|---|---|---|
| 353 | E-Machines e-Tower 366i2 | Celeron/366MHZ (890) | 32MB | 4.3G | $388 | 1,912 |
| 439 | HP Pavillion 8550c | Celeron/500MHZ MMX (1398) | 96MB | 17.0G | $1,012 | 3,001 |
| 168 | Sony Vaio PCV-L600 | Pentium III/ 500MHZ MMX (1650) | 128MB | 10.8G | $1,830 | 22 |
| 412 | IBM Aptiva 595 | Athlon/600MHZ (2108) | 128MB | 20.4G | $1,614 | 1,964 |

malizations and found that the effects of the choice of normalization on the results were small.

We begin by investigating the identification of the preference parameters. We focus on the preference parameters for CPU and RAM for consumers who bought four particular products chosen so as to represent different areas of the characteristics space (see table 3 for a description of the products). Figures 1 and 2 show simulated random draws from the estimated identified sets for the two different models. Note that since the utility function is linear in the parameters, the identified sets are convex polyhedrons. Apparent deviations from that are due to random noise in the simulation draws.[12]

If both the product space and characteristics space were continuous, then the preference parameters would be point-identified (the identified sets would be singletons). In model 1, set identification comes from discreteness of the product space. Since there are 695 products, one might think that the identified sets would always be small. Figure 1 shows that this is usually the case but not always. In particular, product 168 has relatively wide bounds on the preference for RAM, whereas product 412 has relatively wide bounds on both preference parameters. Presumably these wider bounds are due to a lack of close substitutes for these products. The figure does not show this, but there are also a few products with open-ended (including infinity) identified sets in which the product is the best in any single dimension (e.g., fastest CPU or most RAM). Table 4 categorizes the distribution of the preference bounds by how tight the bounds are for each parameter. For model 1 the vast majority of products allow quite tight estimation of the preference parameters. This implies that the distribution of preference pa-

---

[12] Additionally, while the simulation draws are chosen such that they are uniform over the entire identified set, there is no reason that they must still be uniform when projected into a two-dimensional space.
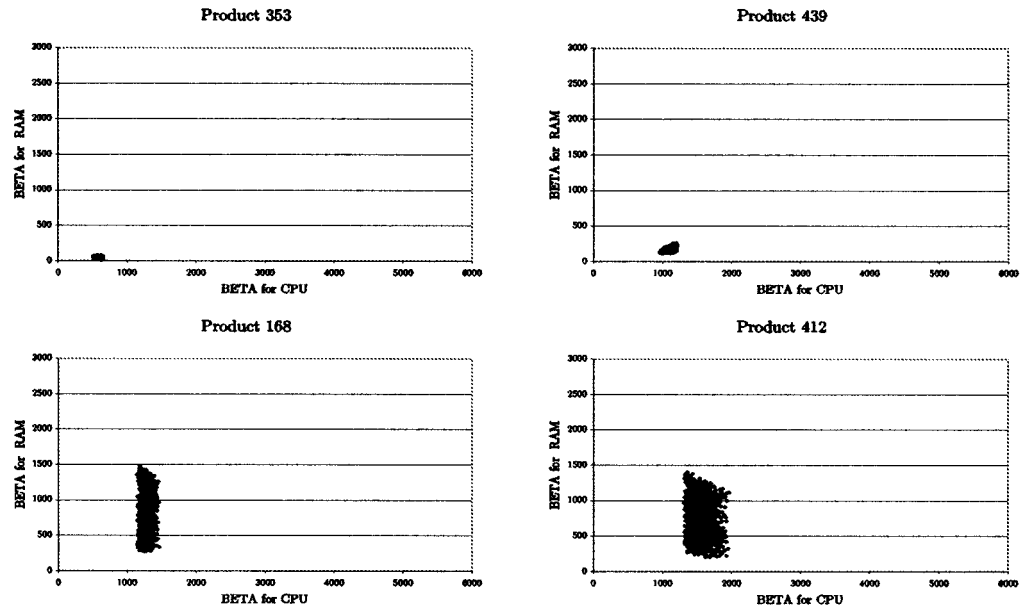
FIG. 1.—Preference parameter identified sets (model 1, continuous characteristics only)

Product 353
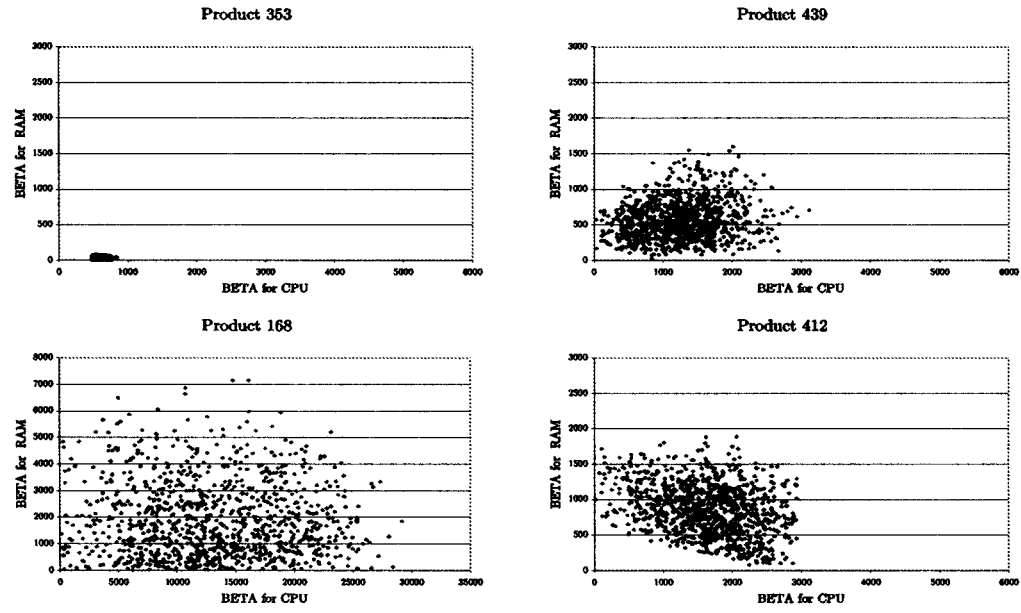
Product 439

Product 168

Product 412

FIG. 2.—Preference parameter identified sets (model 2, all characteristics)

TABLE 4
DISTRIBUTION OF PREFERENCE PARAMETER BOUNDS (Continuous Characteristics)
A. ALL PRODUCTS

|  | % "Small" | % "Medium" | % "Large" | % "Very Large" | % "Huge" |
|---|---|---|---|---|---|
| | Model 1 (Continuous Characteristics) | | | | |
| $\beta_{CPU}$ | 50 | 30 | 10 | 8 | 2 |
| $\beta_{RAM}$ | 50 | 30 | 10 | 7 | 3 |
| $\beta_{HD}$ | 50 | 30 | 10 | 8 | 2 |
| $\beta_{\xi}$ | 50 | 30 | 10 | 8 | 2 |
| | Model 2 (All Characteristics) | | | | |
| $\beta_{CPU}$ | 0 | 7 | 17 | 62 | 13 |
| $\beta_{RAM}$ | 6 | 20 | 54 | 11 | 9 |
| $\beta_{HD}$ | 1 | 14 | 23 | 47 | 15 |
| $\beta_{\xi}$ | 2 | 9 | 24 | 51 | 14 |

B. FOUR REPRESENTATIVE PRODUCTS

| Product | CPU | RAM | Hard Drive | $\xi$ |
|---|---|---|---|---|
| | Model 1 | | | |
| 353 | Small | Small | Small | Medium |
| 439 | Medium | Medium | Large | Large |
| 168 | Medium | Very large | Medium | Large |
| 412 | Very large | Very large | Large | Medium |
| | Model 2 | | | |
| 353 | Large | Small | Large | Very large |
| 439 | Very large | Very large | Huge | Very large |
| 168 | Huge | Huge | Huge | Huge |
| 412 | Very large | Very large | Very large | Very large |

C. DEFINITIONS

|  | Small | Medium | Large | Very Large | Huge |
|---|---|---|---|---|---|
| CPU | <176 | <352 | <642 | <2,500 | >2,500 |
| RAM | <134 | <252 | <1,201 | <2,000 | >2,000 |
| Hard drive | <63 | <137 | <160 | <2,250 | >2,250 |
| $\xi$ | <274 | <739 | <1,292 | <8,000 | >8,000 |

rameters is also very well identified. About 2 percent, or 10 products, have "huge" or open-ended bounds that are not very informative.

Allowing for the remaining discrete characteristics generally widens the bounds considerably. Table 4 shows that the CPU, hard drive, and $\xi$ coefficients now have "very large" bounds for the majority of products, and 10–15 percent of products have "huge" or fairly uninformative bounds. Only the RAM coefficient continues to be identified relatively well. Figure 2 graphs the bounds for the four products referenced above. Product 168 is an example in which the bounds are particularly wide. The reason turns out to be that product 168 has a set of discrete characteristics that is relatively unique, rationalizing purchase of product
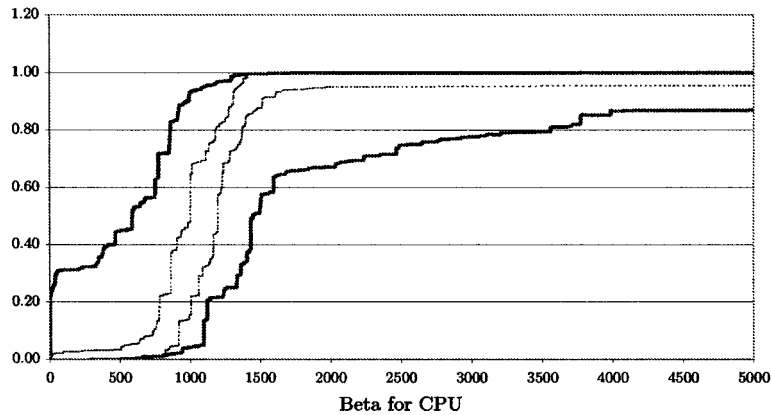
Fig. 3.—Bounds on the distribution function: taste for CPU

168 for a wide range of preferences for CPU and RAM. Finally, note that we are demanding a lot of the data to identify individual preference parameters. Even with these wider bounds, the data are still likely to be quite informative about the distribution of preference parameters as a whole such that it is likely that it could still be estimated quite flexibly.

Putting together the bounds on individual preference parameters allows us to construct bounds on the distribution of preferences in the population. Figure 3 shows the bounds for the distribution of the CPU taste coefficient (results for the other coefficients are similar), where in the figure the dotted lines represent the bounds from model 1 and solid lines represent the bounds from model 2. The graph shows that in model 1 the taste distributions are quite well identified. This result suggests that in these data, discreteness of the product space (the fact that not every point in characteristics space is available) is not very important for identification. The fact that there are a large number of products means that consumers are always quite close to their optimum, which allows precise identification of their preference parameters. An implication of this is that we would have obtained similar results had we ignored the discreteness altogether in model 1 and instead used the first-order conditions for the consumer optimum to infer each consumer's taste coefficients. On the other hand, the bounds are considerably wider for model 2, likely because of both the large number of characteristics in model 2 and the many discrete characteristics. In model 2, discreteness clearly matters for identification.

Since the demand curve represents marginal willingness to pay at each quantity, wider bounds on the preference parameters are likely to lead to wider bounds on the demand curve. This intuition is borne out

in figure 4. In the figure, dotted lines represent the bounds on demand computed using model 1, whereas the solid lines represent the bounds computed using model 2. Note that the models are not nested, so there is no reason for the bounds to coincide or nest. Model 1 generally leads to fairly tight bounds on the demand curve and very elastic demand, perhaps unreasonably so. Model 2 leads to wider bounds but more reasonable "elasticities." (We use the term loosely here to refer to the general slope of demand.) This result is intuitive. In model 1, the product space is only four-dimensional and the characteristics are continuous. Therefore, products appear to be very closely related to each other, resulting in more elastic demand curves. In model 2 the product space is 19-dimensional, and products are, on average, more differentiated. For model 2, implied "elasticities" are approximately in the range of about five to 100 for most products. This is quite high, but perhaps not unreasonable given the very detailed definition of a product here. Such high elasticities would imply very competitive pricing if the market consisted of single-product firms; but in our data several firms have as many as 50–100 products each, so that is not necessarily the case here. These results suggest that, while including the discrete characteristics leads to less precise identification of the model, they seem to be economically important.

Next we investigate the implications of the preference parameter bounds for welfare calculations. First we compute bounds on the welfare loss to individuals if they were to face a different choice set consisting of only the products and prevailing prices available in the first data period (August 1997; see table 5). This experiment leads to losses of 11–16 percent in model 1 and 9–42 percent in model 2. This welfare calculation is similar to those undertaken in constructing price indexes. The welfare bounds imply bounds on the price index (August 1997 = 100) of 54–73 for model 1 and 23–60 for model 2. These bounds are fairly wide, but that is not out of the ordinary in this context.

Next we computed bounds on the welfare loss to each individual if his preferred product were removed from the choice set. This is a calculation similar to those performed by Trajtenberg (1989) and Petrin (2002). Table 6 shows the distribution of the bounds for both models. In model 1, consumers lose very little if their favorite product is removed, with the average maximum loss across products being $15 (this figure excludes the handful of products for which the upper bound is infinite). In model 2, this rises to an average maximum loss of $213, though the median maximum loss is only $59. In our opinion the numbers from model 2 are quite reasonable. In a market in which there are so many similar products available, it is often possible for consumers to switch to a nearly identical product. In this second example, even though
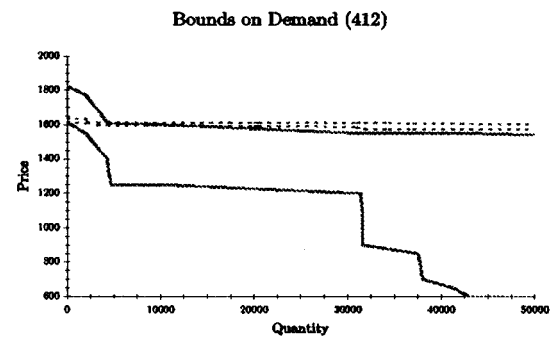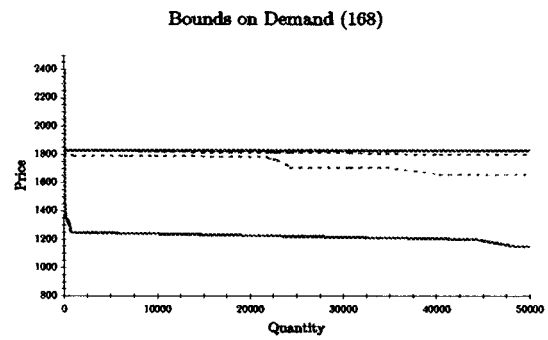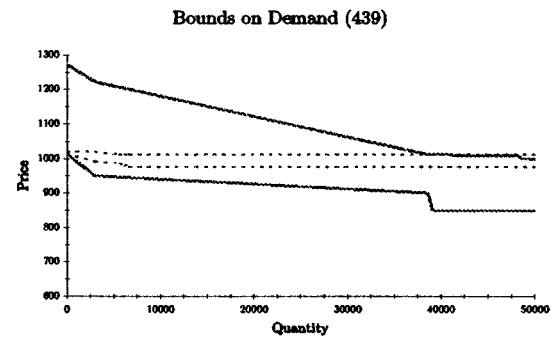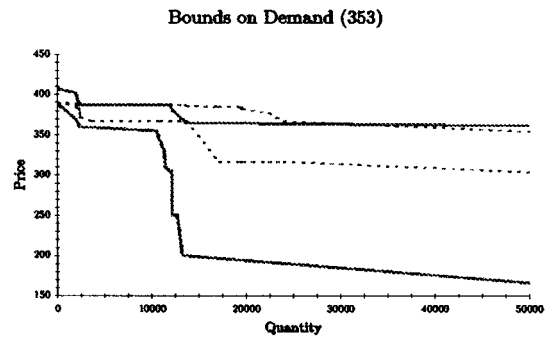
FIG. 4.—Bounds on the demand curve, both models

TABLE 5
DISTRIBUTION OF BOUNDS ON WELFARE LOSS FROM FACING AUGUST 1997 CHOICE SET (%)

|  | PERCENTILE | | | | | | | AVERAGE (Excluding ∞'s) |
|  | .05 | .10 | .25 | .5 | .75 | .9 | .95 |  |
|---|---|---|---|---|---|---|---|---|
| Model 1 | | | | | | | | |
| Lower bound | 7 | 9 | 10 | 11 | 12 | 13 | 14 | 11 |
| Upper bound | 10 | 11 | 12 | 13 | 15 | 19 | 26 | 16 |
| Model 2 | | | | | | | | |
| Lower bound | −7 | −2 | 4 | 11 | 15 | 17 | 19 | 9 |
| Upper bound | 21 | 23 | 25 | 31 | 43 | 67 | 94 | 42 |

TABLE 6
DISTRIBUTION OF BOUNDS ON WELFARE LOSS FROM REMOVING EACH PRODUCT ($)

|  | PERCENTILE | | | | | | | AVERAGE (Excluding ∞'s) |
|  | .05 | .10 | .25 | .5 | .75 | .9 | .95 |  |
|---|---|---|---|---|---|---|---|---|
| Model 1 | | | | | | | | |
| Lower bound | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Upper bound | 0 | 0 | 1 | 3 | 9 | 26 | 64 | 15 |
| Model 2 | | | | | | | | |
| Lower bound | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Upper bound | 6 | 13 | 27 | 59 | 177 | 526 | 851 | 213 |

the preference parameter bounds are quite wide, the welfare bounds end up being quite useful.

## VII.    Conclusions

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) approach in several ways: (1) we allow for imperfect competition, (2) we allow for unobserved product characteristics, (3) we generalize the first-stage estimation to be nonparametric and nonseparable, and (4) we take an alternative approach to the second stage that avoids the recent criticisms of Rosen's second stage and allows for a discrete product space and discrete characteristics. Our hope is that these generalizations will make it much easier to apply hedonic demand models in empirical work in a wider set of applications in the future.

In application of the model to computer demand, allowing for unobserved characteristics was shown to be important because it allowed

rationalization of demand using standard models of utility. Under the assumption that all product characteristics are perfectly observable, as in Rosen (1974), many products that were observed to have positive demand in the data would have been dominated.

We also found that, while the preference parameters for the continuous characteristics were identified quite well in the absence of the discrete characteristics, the presence of the discrete characteristics loosened this identification somewhat. This lack of identification led to wide bounds in some cases (demand, price indexes), but still led to useful bounds on the welfare loss to individuals if their favorite product was removed from the choice set. In cases in which the bounds are wide, it would be possible to make assumptions about the parametric form of the aggregate distribution of preferences in order to get more precise estimates. However, the fact that the bounds were wide shows that results obtained in this way might be sensitive to these additional distributional assumptions.

Finally, while the main disadvantage of our estimation approach is likely to be the first-stage data requirement, this problem did not show up in our data, which are fairly rich and contain a large number of products. However, the first-stage estimates are likely to be somewhat less precise in markets with fewer products.

## Appendix

### Proof of Theorem 1

Part 1: Suppose that $p_{jt} > p_{j't}$ for some market $t$ in which both $j$ and $j'$ have positive demand. Then since $u_i$ is strictly increasing in $c$, $u_i(\boldsymbol{x}_j, \xi_j, y_i - p_{jt}) < u_i(\boldsymbol{x}_{j'}, \xi_{j'}, y_i - p_{j't})$ for all individuals $i$. This implies that demand for $j$ is zero in market $t$, which is a contradiction.

Part 2: Suppose that $p_{jt} \leq p_{j't}$ for some market $t$ in which both $j$ and $j'$ have positive demand. Then since $u_i$ is strictly increasing in $c$ and strictly increasing in $\xi$, $u_i(\boldsymbol{x}_j, \xi_j, y_i - p_{jt}) > u_i(\boldsymbol{x}_{j'}, \xi_{j'}, y_i - p_{j't})$ for all individuals. This implies that demand for $j'$ is zero in market $t$, which is a contradiction.

Part 3: Consider any two products $j$ and $j'$. If $j$ and $j'$ have the same characteristics, then the result holds by part 1 above. Suppose that $j$ and $j'$ have different characteristics in at least one dimension and assume without loss of generality that $p_{jt} > p_{j't}$. Since $u_i$ is Lipschitz continuous in $(\boldsymbol{x}_j, \xi_j)$, we have that

$$|u_i(\boldsymbol{x}_j, \xi_j, y_i - p_{jt}) - u_i(\boldsymbol{x}_{j'}, \xi_{j'}, y_i - p_{jt})| \leq M_1(|\boldsymbol{x}_j - \boldsymbol{x}_{j'}| + |\xi_j - \xi_{j'}|). \qquad \text{(A1)}$$

By a mean value expansion, for all individuals,

$$u_i(\boldsymbol{x}_{j'}, \xi_{j'}, y_i - p_{j't}) = u_i(\boldsymbol{x}_{j'}, \xi_{j'}, y_i - p_{jt}) + (p_{jt} - p_{j't}) \frac{\partial u_i(\boldsymbol{x}_{j'}, \xi_{j'}, y_i - p_{jt}^*)}{\partial c}, \qquad \text{(A2)}$$

where $p_{jt}^* \in [p_{jt}, p_{j't}]$ and varies for each $i$. Plugging (A2) into (A1) gives

$$\left| [u_i(\mathbf{x}_j, \xi_j, y_i - p_{jt}) - u_i(\mathbf{x}_{j'}, \xi_{j'}, y_i - p_{j't})] + (p_{jt} - p_{j't}) \frac{\partial u_i(\mathbf{x}_{j'}, \xi_{j'}, y_i - p_{jt}^*)}{\partial c} \right| \leq$$

$$M_1(|\mathbf{x}_j - \mathbf{x}_{j'}| + |\xi_j - \xi_{j'}|). \tag{A3}$$

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

$$(p_{jt} - p_{j't}) \frac{\partial u_i(\mathbf{x}_{j'}, \xi_{j'}, y_i - p_{jt}^*)}{\partial c} \leq M_1(|\mathbf{x}_j - \mathbf{x}_{j'}| + |\xi_j - \xi_{j'}|)$$

for $i$ that prefer $j$ to $j'$,

$$p_{jt} - p_{j't} \leq \left[ M \left| \frac{\partial u_i(\mathbf{x}_{j'}, \xi_{j'}, y_i - p_{jt}^*)}{\partial c} \right| \right] (|\mathbf{x}_j - \mathbf{x}_{j'}| + |\xi_j - \xi_{j'}|)$$

for $i$ that prefer $j$ to $j'$.

$$\leq \frac{M_1}{\epsilon} (|\mathbf{x}_j - \mathbf{x}_{j'}| + |\xi_j - \xi_{j'}|)$$

$$= M_2(|\mathbf{x}_j - \mathbf{x}_{j'}| + |\xi_j - \xi_{j'}|). \tag{A4}$$

Expression (A4) uses the fact that both products have positive demand to place a limit on how much their prices can vary. The closer the products are in characteristics space, the less their prices can vary without sending demand for the higher-priced product to zero.

### References

Ackerberg, Daniel A., and Marc Rysman. Forthcoming. "Unobserved Product Differentiation in Discrete Choice Models: Estimating Price Elasticities and Welfare Effects." *Rand J. Econ.*

Bajari, Patrick, and Matthew E. Kahn. 2005. "Estimating Housing Demand with an Application to Explaining Racial Segregation in Cities." *J. Bus. and Econ. Statis.* 23 (January): 20–33.

Bartik, Timothy J. 1987. "The Estimation of Demand Parameters in Hedonic Price Models." *J.P.E.* 95 (February): 81–88.

Bayer, Patrick, Robert McMillan, and Kim Rueben. 2004. "An Equilibrium Model of Sorting in an Urban Housing Market: A Study of the Causes and Consequences of Racial Segregation." Working Paper no. 10865 (November), NBER, Cambridge, MA.

Benkard, C. Lanier, and Patrick Bajari. 2004. "Demand Estimation with Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach." Working Paper no. 10278 (February), NBER, Cambridge, MA.

———. 2005. "Hedonic Price Indexes with Unobserved Product Characteristics, and Application to Personal Computers." *J. Bus. and Econ. Statis.* 23 (January): 61–75.

Berry, Steven. 1994. "Estimating Discrete-Choice Models of Product Differentiation." *Rand J. Econ.* 25 (Summer): 242–62.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica* 63 (July): 841–90.

Berry, Steven, and Ariel Pakes. 2001. "Estimating the Pure Hedonic Discrete Choice Model." Manuscript, Yale Univ.

Brown, James N., and Harvey S. Rosen. 1982. "On the Estimation of Structural Hedonic Price Models." *Econometrica* 50 (May): 765–68.

Chernozhukov, Victor, and Christian Hansen. 2005. "An IV Model of Quantile Treatment Effects." *Econometrica* 73 (January): 245–61.

Chernozhukov, Victor, Guido W. Imbens, and Whitney K. Newey. 2004. "Instrumental Variable Identification and Estimation of Nonseparable Models via Quantile Conditions." Manuscript, Massachusetts Inst. Tech.

Chesher, Andrew. 2003. "Identification in Nonseparable Models." *Econometrica* 71 (September): 1405–41.

Court, A. T. 1939. "Hedonic Price Indexes with Automotive Examples." In *The Dynamics of Automobile Demand.* New York: General Motors.

Ekeland, Ivar, James J. Heckman, and Lars Nesheim. 2004. "Identification and Estimation of Hedonic Models." *J.P.E.* 112, no. 1, pt. 2 (February): S60–S109.

Epple, Dennis. 1987. "Hedonic Prices and Implicit Markets: Estimating Demand and Supply Functions for Differentiated Products." *J.P.E.* 95 (February): 59–80.

Fan, Jianqing, and Irène Gijbels. 1996. *Local Polynomial Modelling and Its Applications.* London: Chapman and Hall.

Gale, David, and Hukukane Nikaidô. 1965. "The Jacobian Matrix and Global Univalence of Mappings." *Mathematische Annalen* 159: 81–93.

Goettler, Ronald L., and Ron Shachar. 2001. "Spatial Competition in the Network Television Industry." *Rand J. Econ.* 32 (Winter): 624–56.

Gorman, W. M. 1980. "A Possible Procedure for Analysing Quality Differentials in the Egg Market." *Rev. Econ. Studies* 47 (October): 843–56.

Griliches, Zvi. 1961. "Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change." In *The Price Statistics of the Federal Government: Review, Appraisal, and Recommendations.* New York: NBER.

Haile, Philip A., and Eli Tamer. 2003. "Inference with an Incomplete Model of English Auctions." *J.P.E.* 111 (February): 1–51.

Imbens, Guido W., and Whitney K. Newey. 2003. "Identification and Estimation of Triangular Simultaneous Equations Models without Additivity." Manuscript, Massachusetts Inst. Tech.

Lancaster, Kelvin J. 1966. "A New Approach to Consumer Theory." *J.P.E.* 74 (April): 132–57.

———. 1971. *Consumer Demand: A New Approach.* New York: Columbia Univ. Press.

Manski, Charles F., and John V. Pepper. 2000. "Monotone Instrumental Variables: With an Application to the Returns to Schooling." *Econometrica* 68 (July): 997–1010.

Manski, Charles F., and Elie Tamer. 2002. "Inference on Regressions with Interval Data on a Regressor or Outcome." *Econometrica* 70 (March): 519–46.

Matzkin, Rosa L. 2003. "Nonparametric Estimation of Nonadditive Random Functions." *Econometrica* 71 (September): 1332–75.

Motzkin, T. S., H. Raiffa, G. L. Thompson, and R. M. Thrall. 1953. "The Double Description Method." In *Contributions to the Theory of Games,* vol. 2, edited by H. W. Kuhn and A. W. Tucker. Princeton, NJ: Princeton Univ. Press.

Nevo, Aviv. 2001. "Measuring Market Power in the Ready-to-Eat Cereal Industry." *Econometrica* 69 (March): 307–42.

Pakes, Ariel, Jack Porter, Kate Ho, and Joy Ishii. 2005. "Moment Inequalities and Their Application." Manuscript, Harvard Univ.

Petrin, Amil. 2002. "Quantifying the Benefits of New Products: The Case of the Minivan." *J.P.E.* 110 (August): 705–29.

Petrin, Amil, and Kenneth Train. 2004. "Omitted Product Attributes in Discrete Choice Models." Manuscript, Univ. Chicago.

Rosen, Sherwin. 1974. "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition." *J.P.E.* 82 (January/February): 34–55.

Rossi, Peter E., Robert E. McCulloch, and Greg M. Allenby. 1996. "The Value of Purchase History Data in Target Marketing." *Marketing Sci.* 15 (4): 321–40.

Trajtenberg, Manuel. 1989. "The Welfare Analysis of Product Innovations, with an Application to Computed Tomography Scanners." *J.P.E.* 97 (April): 444–79.