Discrete Choice Models as Structural Models of Demand: Some Economic Implications of Common Approaches*

Patrick Bajari†
Stanford University and NBER

C. Lanier Benkard‡
Stanford University and NBER

March 2003

Abstract

We derive several properties of commonly used discrete choice models that are potentially undesirable if these models are to be used as structural models of demand. Specifically, we show that as the number of goods in the market becomes large in these models, i) no product has a perfect substitute, ii) Bertrand-Nash markups do not converge to zero, iii) the fraction of utility derived from the observed product characteristics tends toward zero, iv) if an outside good is present, then the deterministic portion of utility from the inside goods must tend to negative infinity and v) there is always some consumer that is willing to pay an arbitrarily large sum for each good in the market. These results support further research on alternative demand models such as those of Berry and Pakes (2001), Ackerberg and Rysman (2001), or Bajari and Benkard (2003).

*We thank Daniel Ackerberg, Steven Berry, Richard Blundell, Jonathon Levin, Peter Reiss, and Ed Vytlacil for many helpful comments on an earlier draft of this paper, as well as seminar participants at Carnegie Mellon, Northwestern, Stanford, UBC, UCL, UCLA, Wash. Univ. in St. Louis, Wisconsin, and Yale. We gratefully acknowledge financial support of NSF grant #SES-0112106. Any remaining errors are our own.

†Dept. of Economics, Stanford University, Stanford, CA 94305, bajari@leland.stanford.edu

‡Graduate School of Business, Stanford University, Stanford, CA 94305-5015, lanierb@stanford.edu

1 Introduction.

In empirical studies of differentiated product markets, it is common practice to estimate random utility models (RUM) of consumer demand. In a RUM, the consumer's utility is a function of product characteristics, a composite commodity and a stochastic, household specific, product level taste shock. In many markets, such as autos, computers or housing, there are thousands, if not tens of thousands of unique products available to consumers. As a result, each household's utility function contains a very large number of these random taste shocks. Since households choose the good with the maximum utility, it is important to pay attention to how demand models are affected by large draws of the taste shocks.

In general, RUMs provide a quite flexible framework for modeling consumer preferences. However, in practice, in order to facilitate estimation of RUMs, practitioners typically make restrictive functional form assumptions about the distribution of the taste shocks. For example, the logit model is attractive from a computational point of view because the likelihood function can be evaluated in closed form. The multinomial probit model can be studied in a computationally efficient manner using simulation.

While commonly used RUM models greatly simplify the computational burden of estimation, some recent papers have found that such simplifications can also lead to implausible results. For example, Petrin (2002) studies the problem of measuring the welfare benefits from the introduction of the minivan. In his logit estimates, he finds that the average compensating variation for minivan purchasers is $7,414, for a vehicle that sold for just $8,722, with 10% of consumers willing to pay over $20,000 for the option to purchase a minivan. Even his random coefficients logit estimates have several percent of consumers willing to pay over $20,000 for this option. Similarly, in Ackerberg and Rysman's (2001) study of the demand for phone books, the optimal number of products goes from greater than 10 to just 5 when a model with a logit error term is used versus a model that uses a more general error structure. In both cases, the authors found that their results were being generated from large realizations of the logit error term in the demand system.

In this paper, we investigate the economic implications of common assumptions on the error term in RUMs. Many RUMs have the following properties: 1) the model includes an additive error term whose conditional support is unbounded, 2) the deterministic part of the utility function satisfies standard continuity and monotonicity conditions, and 3) the hazard rate of the error distribution is bounded above. We show that these assumptions can lead to several potentially unappealing features of the demand model.

First, demand for every product is positive for all sets of prices. This implies that some consumers are unwilling to substitute away from their preferred product at any price. For many products, this property is a priori unreasonable.

Second, when the number of products becomes large, the share of the outside good must go to zero, meaning that every consumer purchases some variety of the good. In many markets this property is also a priori unreasonable. Similarly, if the share of the outside good can always be bounded away from zero, then the deterministic part of utility from the inside goods for all of the consumers who purchase the outside good must tend to negative infinity as the number of products becomes large, suggesting that the parameters of the model are not stable to changes in the number of goods.

Third, in Generalized Extreme Value (GEV) based models, even if the number of products becomes infinite, no product has a perfect substitute. That is, each individual would almost surely suffer a utility loss bounded away from zero if she was forced to consume a product other than her preferred alternative. This property shows that the product space can never become crowded as implied by standard models of horizontal and vertical product differentiation with a continuum of goods.

Fourth, in GEV based models, as the number of products becomes large, a Bertrand-Nash equilibrium does not tend toward the perfectly competitive outcome where all firms price at marginal cost. This shows that RUMs always build in "excess" market power for the firms. As a result, some RUMs may generate misleading implications if they are used, for example, in merger analysis. Note that the last two properties do not hold for probit models, suggesting that when there are a large number of products, the probit model may be preferable to the logit model in some applications.

Fifth, as the number of products becomes large, the ratio of the error term to the deterministic part of utility at the chosen product becomes one. That is, all of the consumer's utility is derived from the random error term. This suggests that RUMs may tend to understate the consumers' welfare from the observed product characteristics.

Sixth, as the number of products becomes large, the compensating variation for removing all of the inside goods tends to infinity for each individual. This last result is consistent with the empirical findings of Petrin (2002) and Ackerberg and Rysman (2001).

We are not the first to raise many of these issues. For instance, Ackerberg and Rysman (2001), Berry and Pakes (2000), Caplin and Nalebuff (1991), and Petrin (2002) note that the logit model tends to overstate the benefits from product variety. Also, Andersen, de Palma and Thisse (1992) establish that Bertrand competition does not converge to perfect competition in the logit model. However, to the best of our knowledge, our second, third and fifth undesirable property listed above are new to the literature. In addition, we believe that the result that Bertrand competition does converge to perfect competition for the probit model is new. For the other properties, we formalize and generalize previous work, establishing exactly which assumptions are at fault in each case.

One possible response to our criticisms is that many of these undesireable properties can be mitigated during estimation of the model. For example, for any given set of parameters the share of the outside good goes to zero as you add products to the market. However, if there are a large number of goods in the market under study, then during estimation the parameter estimates will adjust the mean utility of inside goods downward until the correct share of the outside good is obtained. This mitigates the undesireable properties of the model with respect to the model's ability to match the data. However, the point of this paper is to evaluate the demand system as a structural model in which the parameters are primitives of the model and thus are fixed. In any policy experiment, the parameters would necessarily be held fixed, and the properties listed above are therefore likely to impact the results. Furthermore, it seems strange to think about consistency of parameter estimates in a model which (in some cases) requires the number of products to become large for consistency to hold, but in which we know that adding products must necessarily lead to a change in the parameters.

In practice, these properties are likely to be more problematic in some empirical applications than others. We think that they are more likely to pose a problem if the number of goods in the market of interest changes in the data used to estimate the model, or in counterfactuals involving changes in the number of goods. For example, consumer surplus calculations, price indexes, and applications on entry and exit, are counterfactual experiments that involve changing the number of goods. Of course, the extent of the problem is an empirical question. We do not attempt to address that here other than through the references mentioned above.

Our results lend some support to further research on models in which willingness to pay estimates do not depend so heavily on realizations of the error term in the RUM, especially when the number of products is large. For example, Berry and Pakes (2001) and Bajari and Benkard (2003) develop discrete choice models with heterogeneity in the utility function, but which do not have a random error term. Alternatively, Ackerberg and Rysman (2001), propose allowing the mean or the variance of the logit error term to change with the number of products in the market.

Note that there are also some cases where a random error term may be desirable. For example, Bajari and Benkard (2003) point out that the random error term may be a good way of modeling consumers' imperfect information about the choice set. In a model that has no random error term, they find that the assumption that consumers have perfect information about all products leads to estimates of price elasticities that are counterintuitively high. However, if the random error term is being used to model imperfect information, then its treatment in welfare calculations would be different than the standard approach (dating back at least to Domencich and McFadden (1975)), and many of the results of this paper would not necessarily hold.

2 The Model

In this section, we develop a fairly general semi-parametric discrete choice model. This model nests as special cases many commonly used RUMs such as the logit, nested logit, GEV, multinomial probit, as well as random coefficients versions of these models, such as BLP models.

In the model, each consumer chooses between J mutually exclusive alternatives. We index consumers by $i \in 1..I$ and products by $j \in 1..J$. Following the previous literature, we assume that individuals' utility functions can be written as a function of individual characteristics (describing individual tastes), product characteristics, and an additively separable random error term:

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

In equation (1), $x_j \equiv (x_{j,1}, x_{j,2}, ..., x_{j,K})$ is a K-dimensional vector of characteristics associated with product j. We assume that $x \in \mathcal{X}$, where $\mathcal{X} \subseteq R^K$ is a compact set. In addition, $p_j \in R^+$ is the price of product j, $y_i \in R^+$ is the income of consumer i, β_i is a vector of individual taste parameters with support $\mathcal{B} \subseteq R^B$, and ϵ_{ij} is an individual and product specific random error term. The term $y_i - p_j$ represents consumption of all other goods, which we treat as a composite commodity denoted as c. Therefore the utility function in (1) should be thought of as a direct utility function with preferences over the characteristics of inside goods and the composite outside good, with the budget constraint substituted in.[1] The function $u(\cdot)$ is assumed to take a known parametric form that is constant across individuals.

We assume that the utility obtained from not purchasing any variety of the good is also a function of a random error term,

$$u_{i0} = u(\tilde{0}, y_i, \beta_i) + \epsilon_{i0}. \tag{2}$$

It is not necessary to include the outside good in the model for most of what follows. We include it because its presence underscores some of the undesirable properties of the model,

[1] The price of the composite commodity is normalized to one.

and because much of the previous literature models the outside good similarly.

In the model, consumers are rational utility maximizers. Consumer i chooses product j if and only if j maximizes utility,

$$i \text{ chooses } j \iff u_{ij} \geq u_{ik} \text{ for all } k \neq j. \tag{3}$$

Manski (1977) suggests that there are four sources of uncertainty that justify the use of the random error term in the model (1)-(5):

1. **Unobserved product characteristics.** The vector x_j may not include all product characteristics that enter into the consumer's utility function.
2. **Unobserved consumer heterogeneity.** The distribution of consumer tastes may differ in the population in ways that cannot be explained by income or other available demographic information.
3. **Measurement error.** The values of the x_j or p_j may be mismeasured by the economist.
4. **Functional misspecification.** The econometrician typically does not know the true functional form for u_{ij}.

Let $s_{i,j} \equiv P_i(j|\beta_i, y_i)$ denote the probability that consumer i chooses product j conditional on β_i and y_i, and let $\epsilon_{i,-j}$ denote the vector of error terms for individual i excluding product j. By equation (3) it follows that:

$$P_i(j|\beta_i, y_i) = \int_{-\infty}^{\infty} \int_{-\infty}^{u_{i,j}-u_{i,0}} \cdots \int_{-\infty}^{u_{i,j}-u_{i,j-1}} \int_{-\infty}^{u_{i,j}-u_{i,j+1}} \cdots \int_{-\infty}^{u_{i,j}-u_{i,J}} f(\epsilon|\beta_i, y_i) d\epsilon_{i,-j} d\epsilon_{i,j}. \tag{4}$$

That is, the probability that consumer i chooses product j is the probability that the realization of ϵ makes choice j utility maximizing for consumer i. If the researcher has access to micro data containing individual level choices, equation (4) can be used to construct a likelihood function.

Let $s_j \equiv P(j)$ denote the probability that j is chosen averaging over the $i = 1, ..., I$ consumers. Then,

$$P(j) = \int P_i(j|\beta_i, y_i) dF(\beta_i, y_i). \tag{5}$$

In equation (5) we integrate out over the population distribution of β and y in order to compute the probability that product j is chosen by a randomly selected consumer. In product markets, $P(j)$ is typically interpreted as the demand for product j. If only aggregate market shares are observed, then equation (5) can be used to construct a likelihood function. The object of interest is $F(\beta, y, \epsilon)$, which is typically estimated with some parametric and independence restrictions. We discuss these restrictions in detail in the next section.

3 Economic Implications of Standard Discrete Choice Models.

In its general form, the model in equations (1)-(5) is quite flexible. It nests as a special case many deterministic discrete choice models (e.g. Berry and Pakes (2001), Bajari and Benkard (2001)) as well as all commonly used econometric models such as logit, GEV, probit, and BLP. However, in applied work, the model is typically not estimated in full generality due to the complexity of computing the integral (4). Instead, econometricians typically make restrictive functional form and independence assumptions about the joint distribution $F(\beta, y, \epsilon)$ in order to simplify the computation of (4). For example, in the random coefficients logit model it is assumed that ϵ_{ij} is *iid*, independent of β_i and y_i, and is distributed extreme value. In that case, the integral (4) has a closed form solution. In the random coefficients probit model, it is assumed that ϵ_{ij} is independent of β_i and y_i and normally distributed. In that case, simulation methods such as Gibbs sampling can be used to compute (4).

In this section we show that these commonly used functional form and independence assumptions have implications that are undesirable if the intention is to use the model as a structural model of demand.

3.1 Assumptions

The first assumption we make is a common assumption regarding the independence of the consumer taste coefficients and the error terms.

Assumption I The vector of errors, ε, is independent of consumer tastes and income. That is, $f(\beta, y, \varepsilon) = f_{\beta,y}(\beta, y) f_{\epsilon}(\epsilon)$.

This assumption is made for convenience only, while it may be unpalatable in many applications, it is not important to this paper. All of the results that follow can be shown to hold without it.

Next, we restrict the set of utility functions that we will consider.

Assumption U

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

(ii) For every $(\beta, y) \in B \times R^{+}$ and every $0 < p < y$, $|u(., y - p, \beta)| < \infty$.

Assumption $U(i)$ says that the deterministic part of the utility functions is continuous and that individuals have monotone preferences with respect to the composite commodity. Assumption $U(ii)$ says that the utility function as defined over characteristics is bounded for every individual so long as the budget constraint is satisfied. Assumption U holds in all applications of discrete choice in the previous literature that we are aware of.

The next assumption is the critical assumption driving our results.

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

Assumption R amounts to assuming that the conditional error distributions have unbounded upper support. All GEV models, the probit model, and all GEV- and probit-based random coefficients models satisfy R, so long as the errors have strictly positive variance, mean that is bounded below, and so long as the errors are not perfectly correlated. Assumption R guarantees that $\lim_{J\to\infty} \max_{j\in 1..J} \epsilon_{ij} = \infty$ a.s.[2]

3.2 Property One: The Shape of the Demand Curve

The first property implied by these assumptions is that the demand curve is never bounded above.

1. Demand is positive for every price vector. Suppose that assumptions I, R and U hold. Suppose further that, either: (i) $u(x, c, \beta)$ is linear in c, or (ii) for all $\bar{\beta} \in B$, $F(y|\bar{\beta})$ has full support on R^+. Then in the model described above, for every product, demand is strictly positive for every price.

Conditions (i) and (ii) are satisfied in much, if not all, of the previous literature on discrete choice. For example, condition (i) is typically satisfied in standard applications of logit and probit. Under condition (i), income does not affect individuals' choices and thus income can be omitted from the analysis. Condition (ii) is satisfied in Berry, Levinsohn, and Pakes (1995) and every application of BLP style models that we are aware of that do not satisfy (i).

This first property shows that standard discrete choice econometric models imply a very particular, and perhaps undesirable, shape for product level demand curves. Because product-level demand curves never touch the vertical (price) axis, and consumer surplus equals the

[2] The proof is a straightforward application of the Borell-Cantelli Lemma. See appendix.

area underneath the demand curve, it seems likely that the model is a priori biased toward generating large amounts of consumer surplus from each product. The model enforces some differentiation across all products, regardless of the similarity of their characteristics.

One could argue that, because only a narrow range of prices are typically observed for each good, it is not possible to estimate the amount of consumer surplus generated by a good without making functional form assumptions for the demand curve. However, this is not entirely true. If products can be well represented in characteristics space, then standard revealed preference arguments would place an upper bound on consumers' willingness to pay for most products in the market. The intuition for this is that a consumer could obtain more of every characteristic by buying another good in the market that has a finite price, thus placing an upper bound on the amount any consumer should be willing to pay for the good in question.[3] These upper bounds would conflict with property one, and thus would reject the shape of the demand curve that results from the models commonly used. See Bajari and Benkard (2003) for a more detailed discussion of revealed preference in characteristics models.

Furthermore, even if it were true that the shape of the demand curve was not identified at high price levels, an unbounded demand curve is likely to be undesirable for welfare measurement. For example, Hausman (1997) uses linear and quadratic approximations to the demand curve in order to make welfare calculations, favoring them over the CES specification, which has an unbounded demand curve.

Property one also implies that two different firms may sell products with the exact same product characteristics, x, at different prices in the same market. By property one, both products would have positive demand. This can happen because the vector of product characteristics, x, is not a complete description of the product in these models due to the error terms. In many markets this property may be unreasonable.

[3] However, for goods that are the best in any single dimension, no upper bound is obtainable.

3.3 Properties of the Model When the Number of Products Increases

We now show that the behavior of existing models is counterintuitive when the number of products in the market becomes large. These results are most directly relevant to applications with large numbers of products. However, we note that these counterintuitive properties also hold loosely when the number of products in the market changes. Thus, we think that these results are also relevant to applications in which the number of products changes a lot in the data, as well as counter-factual policy experiments involving changes in the number of products. Such experiments include price index calculations, measuring the welfare from new products or new inventions, mergers, etc.

Note that in this section we have intentionally omitted the process by which the products are added when the limit is taken because the properties listed hold regardless of what process is generating the added products so long as the assumptions listed are satisfied.

2. **Share of the Outside Good** Let s_{i0} denote the probability that individual i chooses the outside good. If I, R, and U hold, then as $J \to \infty$ either $s_{i0} \to 0$, or $s_{i0} > 0$ and $u(x_j, y_i - p_j, \beta_i) \to -\infty$ for all but a finite set of goods.

If we are trying to describe choice behavior in a narrowly defined market, it seems unreasonable that the share of the outside good tends to zero when there are many varieties of the good. Some characteristics are typically shared by all inside goods (e.g., cell-phones are typically used to make telephone calls, breakfast cereals are typically eaten for breakfast). If an individual has a strong negative taste for a common characteristic of the inside good (e.g., they have no cell-phone service in their area, they do not like cereal, or they do not eat breakfast), then no matter how many varieties are available we should not expect the individual to purchase the good. At very least it would be desirable for the structural demand model to be rich enough to allow for the possibility that the outside good retains positive share in the limit.

Note that this property is not limited to just the outside good. For any fixed set of parameters the model similarly implies that the share of every good goes to zero as products are added to the market, so long as the products added are of sufficiently high mean quality. This property also contradicts the intuition of many theoretical differentiated products models, which suggest that the location where products enter in characteristics space should matter in determining whether or not shares go to zero in the limit.

We now list one additional assumption regarding the error distribution:

Assumption H For each j, the limit as ϵ_j tends to the upper limit of its support of the hazard rate of $F_{\epsilon_j}(\cdot)$ is infinite, i.e., $\lim_{\epsilon \to b} \frac{f_{\epsilon_j}(\epsilon)}{1-F_{\epsilon_j}(\epsilon)} = \infty$, where b is the upper end of the support of $F_{\epsilon_j}(\cdot)$ and b may equal infinity.

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

3. Lack of Perfect Substitutes Suppose that ϵ_{ij} is *iid* and that I, R, and U hold, but H does not hold. Then each product almost surely does not have a perfect substitute even as $J \to \infty$. That is, even when the number of products is infinite, each individual would suffer utility losses that are almost surely bounded away from zero if her first choice product were removed from the choice set.

4. Lack of Perfect Competition Suppose that ϵ_{ij} is *iid* and that I, R, and U hold but H does not hold. Then in a symmetric Bertrand-Nash price setting equilibrium with single product firms, markups are almost surely bounded away from zero when $J \to \infty$.

Properties three and four cover only the *iid* case for simplicity. They are closely related so we discuss them together. Property three implies that, even in the limiting case, the assumptions commonly maintained (e.g., in logit, GEV, and random coefficients logit models) are not sufficient to imply that individuals would be willing to switch to their second favorite product with zero compensation when the number of products becomes large. As a result, markups also remain bounded away from zero in the limit.[4]

Again, if we are considering a narrowly defined market, then we might expect that the product space should fill up eventually and products should become close substitutes in the limit. The functional form assumed in extreme value based models does not allow for this possibility.

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

5. Contribution of Observed Characteristics Suppose that assumptions I, R and U hold. Then the contribution of the observed characteristics to utility almost surely goes to zero as the number of products becomes large. That is, $\lim_{J\to\infty} \frac{\epsilon_{ij^*}}{u_{ij^*}} = 1$ a.s., where $j^* = \arg\max_{j\in 0..J} u_{ij}$.

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

6. Compensating Variation Suppose that assumptions I, R, and U hold. Then as the

[4] Anderson, de Palma, and Thisse also show this for the standard logit model.

number of products becomes large the compensating variation for removing all of the inside goods almost surely tends to infinity for every individual.

Property six singles out a problem with using discrete choice models for welfare analysis. The model implies that with enough products to choose from every individual needs arbitrarily large amounts of income to be as well off with the outside good alone as with the inside good. The implication is that every individual is costlessly receiving arbitrarily large (relative to income or price) levels of utility from something about the good that we cannot observe.

3.4 Properties of the Model in Markets With Large Numbers of Consumers

All of the properties above are driven by properties of the random error term, particularly through changes in its dimension driven by changes in the number of products. Caplin and Nalebuff (1991) show that one interpretation of the error terms in the standard discrete choice econometric model is as a "taste for products", with the following construction:

$$\epsilon_{ij} = \lambda_i' \eta_j. \tag{6}$$

where λ_i is individual i's J-dimensional random vector of tastes for each product, and η_j is a vector of zeros with a one in the jth element. This construction makes it clear that the standard econometric models are special cases of pure characteristics models in which individuals have preferences (with a specific distribution) over product dummies. In this construction, by definition each product is unique, leading to the properties mentioned above.

A large number of goods implies a high dimensional error term and exacerbates these counter-intuitive properties. However, even if the number of products in a particular application is small, it may still be subject to the criticisms above because another way that the dimension of the error term can become large is through the market size.

In the standard discrete choice econometric models, the dimension of the error vector is $I * J$, where I is the number of individuals and J is the number of products. That is, the error vector consists of $I*J$ draws from some distribution. Thus, if just a single product is added to the choice set, the dimension of the error term increases by I, which is typically a very large number. For example, in applications to demand for U.S. households I is on the order of 100 million. In 100 million random draws from any distribution with full support, particularly thick-tailed distributions like the extreme value, large draws can become highly likely. In the structural demand model, these large error draws imply large welfare effects and low substitutability across products. Past empirical work has shown, not surprisingly, that under these conditions the undesirable properties above show up quite strongly in practice even in applications with moderate numbers of products (see, e.g., Petrin (2002)).

4 Alternative Models

In section 3, we showed that many commonly used RUMs have some unappealing economic properties as the number of products becomes large. One way to avoid this set of assumptions is to allow the distribution of the error term to change with the choice set. Such an approach is outlined in detail in Ackerberg and Rysman (2001).[5] Another alternative, which we discuss in more detail here, is to use a "pure hedonic" demand model which elminates the *iid* error term in the utility function. This is the approach of Berry and Pakes (2000) and Bajari and Benkard (2003).

Hedonic models of demand for differentiated products have been used extensively in the past. Examples include models of horizontal product differentiation such as Hotelling (1929), Gorman (1980) and Lancaster (1966), models of vertical product differentiation, such as Shaked and Sutton (1987) and Bresnahan (1987), as well as Rosen's (1974) model.

In the pure hedonic model, commodities are fully described by a collection of a finite number of attributes, x_j. Each consumer i has a utility function $u_i(x_j, c)$ and she chooses a product $j \in J$ along with a composite commodity $c \in R_+$ that maximizes utility. Let the price

[5] Ackerberg and Rysman allow the mean or variance of the error term to fall as more products are added to the market. Their model does not necessarily satisfy our assumption R, and thus does not have the properties listed above.

of commodity j be p_j and normalize the price of the composite commodity to one. Then consumer i's utility maximization problem can be written:

$$\max_{(j,c)} \; u_i(x_j, \xi_j, c) \tag{7}$$

$$\text{s.t.} \qquad p_j + c \leq y_i. \tag{8}$$

As an example of the pure hedonic model, suppose that each product has exactly 3 characteristics and that consumers have preferences as in the equation below:

$$u_{ij} = \beta_{1,i} x_{j,1} + \beta_{2,i} x_{j,2} + \beta_{3,i} x_{j,3} + (y_i - p_j). \tag{9}$$

Note that in (9) we assume that an interior solution exists, which justifies substituting the budget constraint into the utility function. In addition, the coefficient on $(y_i - p_j)$ is normalized to one for each individual without loss of generality.

In the model (9), each consumer i has a unique taste coefficient for each characteristic. In Bajari and Benkard (2003), we show that it is possible under certain assumptions to identify consumer i's taste coefficients from standard data sets that include price, quantities and product characteristics. Bajari and Benkard (2003) also show that the taste coefficients, β_i will typically be just identified if there are as many taste coefficients as product characteristics.

The hedonic model of Bajari and Benkard (2003) is similar to many commonly used RUMs except for the following two features:

1. There is no random idiosyncratic taste shock ε_{ij}.
2. No parametric or independence restrictions are imposed on the joint distribution of β_i.

It can easily be seen that the hedonic model does not impose many of the undesirable assumptions of standard discrete choice models. First, in the hedonic model, it is not always the case that demand is positive at any price. Consider the demand for product j, suppose that there exists a product j' such that:

$$x_{j',k} > x_{j,k} \text{ for } k = 1, ..., K. \tag{10}$$

If $p_{j'} < p_j$ then the demand for product j will be zero since product j' has a higher value of all characteristics but has a lower price.

Furthermore, for individuals with low preference for characteristics of the inside good relative to their preference for the composite commodity, only a very low price will induce them to purchase the good. If a consumer's willingness to pay for characteristics of the good is below the marginal cost of production, then it may be that no rational price is low enough to induce purchase. Thus, the share of the outside good does not necessarily tend to zero as more products enter the market.

If the distance between the characteristics of product j and j' is small and preferences are Lipschitz continuous, then in the pure hedonic model these products will be close substitutes. As a result, as the number of products becomes infinite, all products will have a perfect substitutes and markups in Bertrand price competition will tend to zero.

Finally, the pure hedonic model does not imply that a continuum of products provides consumers with infinite utility relative to income or price. Thus, the compensating variation for removing all inside goods remains bounded even in that case.

However, the hedonic model also has limitations. In the hedonic model, not all products are required to be strong gross substitutes. The number of cross price elasticities for product j that are strictly greater than zero is a function of the number of product characteristics. This is potentially unappealing in markets where there are only a handful of observed characteristics.

In addition, the empirical results in Bajari and Benkard (2003) suggest that if many products are included in the choice set, then the perfect information assumption in the pure hedonic model can lead to demand curves that are too elastic. To summarize their results, one reason for including the random error term in the utility function is that it could represent imperfect information (e.g. due to a cost of acquiring information about products). Leaving out this imperfect information may imply too high a degree of substituteability across products.

5 Conclusions.

In this paper we have shown that standard discrete choice models have some undesirable economic properties when viewed as structural models of demand. These properties are primarily driven by functional form assumptions about the random error term introduced into the model for estimation purposes. They hold not only in the logit model, but in any RUM with the following three properties: 1) the conditional support of the error term is unbounded, 2) the deterministic part of the utility function satisfies standard continuity and monotonicity conditions, and 3) the hazard rate of the error term is bounded above. Due to the computational complexity of estimating RUMs, properties (1)-(3) are maintained in most applications we are aware of.

We cautiously conclude that these results support further research on alternative demand models. For example, the random utility framework of Ackerberg and Rysman (2001), and the pure hedonic model of Berry and Pakes (2000) and Bajari and Benkard (2003), do not necessarily have the undesirable properties derived in section 3. Aggregate demand models such as those used in Hausman (1997) also do not necessarily have these properties. We speculate that there may also be alternative RUM models that eliminate these properties. However, that is an area for further research.

6 Appendix

6.1 Proofs for Section 2

6.1.1 Proof of Property 1

Consider demand at any point $(x_j, p_j, x_{-j}, p_{-j})$. In the case of (i), we fix $\bar{\beta}_i$ arbitrarily and choose $\bar{y}_i$ such that $F(\bar{y}_i|\bar{\beta}) > 0$. In the case of (ii), we fix $\bar{\beta}_i \in B$ arbitrarily and choose $\bar{y}_i$ such that $F(\bar{y}_i|\bar{\beta}) > 0$ and $\bar{y}_i > p_j$. This can be done since under (ii) y_i has full support conditional on β. Set $\epsilon_{ik} = 0$ for all $k \neq j$. Conditional on these values, product j is preferred to all other products if and only if

$$\epsilon_{ij} > \max_{k \neq j}\{u(x_k, \bar{y}_i - p_k, \bar{\beta}_i)\} - u(x_j, \bar{y}_i - p_j, \bar{\beta}_i) \equiv \bar{u}_k - \bar{u}_j. \tag{11}$$

By R, the probability corresponding to (11) is strictly positive.

Let

$$A_j = \{(y, \beta, \epsilon) \in R_+ \times B \times R^{J+1}| \quad u_{ij} \geq u_{ik} \quad \forall k \in 0..J\} \tag{12}$$

A_j represents the set of consumer demographics, taste coefficients, and error terms that rationalize a consumer choosing choice j. In order to find total demand for product j, we simply integrate A_j over the distribution of unobservables to get market share, and then multiply by the market size, M.

$$q_j(x, p; \theta) = M \int_{R^{J+1}} \int_B \int_{R_+} A_j \qquad dF(\beta, y) dF(\epsilon) \tag{13}$$

Thus, using the same point above,

$$q_j(x_j, p_j, x_{-j}, p_{-j}) > M * Prob[\epsilon_{ij} > \bar{u}_k - \bar{u}_j|\bar{\beta}, \bar{y}_i] f(\bar{\beta}, \bar{y}_i) > 0 \tag{14}$$

Since we chose the vector of prices arbitrarily, demand is positive for every good for every price vector.

6.1.2 Proof of Properties 2 and 5

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

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

$$Pr(A_n) = Pr(\epsilon_{i0} < M)Pr(\epsilon_{i1} < M|\epsilon_{i0} < M) * ... * Pr(\epsilon_{in} < M|\epsilon_{i,-n} < M) \quad (15)$$

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

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

6.1.3 Property 6

By the previous result,

$$\lim_{J\to\infty} \max_{j\in 1..J} \epsilon_{ij} = \infty \quad (16)$$

For individual i, the compensating variation for removing the inside goods, CV, is the solution to,

$$u(\tilde{0}, y_i + CV, \beta_i) = \max_{j\in 1..J}\{u(x_j, y_i - p_j, \beta_i) + \epsilon_{ij}\} - \epsilon_{i0} \quad (17)$$

Because utility is bounded, for any given individual the right hand side tends to infinity with J. (Technically, we also need to assume that products are added in such a way that the number of products that are within consumer i's budget tends to infinity with J.) Since preferences for c are monotone, it must be that CV does too.

6.1.4 Properties 3 and 4

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

We show two proofs: 1) $\lim_{J\to\infty} E[\epsilon_1^J - \epsilon_2^J] = 0$, where ϵ_1^J is the highest of J draws on ϵ and ϵ_2^J is the second highest, if and only if H holds; 2) as the number of products becomes large the markup in a symmetric Bertrand-Nash price-setting equilibrium with single product firms tends to 0 if and only if H holds.

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

2. Consider J identical single product firms facing a demand system generated by a discrete choice model where the utility function is $u_{ij} = p_j - \epsilon_{ij}$ and the errors are assumed to be *iid*. In a symmetric Bertrand-Nash price setting equilibrium, all firms' prices are the same and

each firm has equal market share $s_j = 1/J$. The markup is $\frac{s_j}{-\frac{\partial s_j}{\partial p_j}}$. We now consider $\frac{\partial s_j}{\partial p_j}$:

$$
\begin{aligned}
s_j &= Pr(\epsilon_k \leq \epsilon_j + p_k - p_j \forall k \neq j) && (18)\\
&= \int_{-\infty}^{\infty} Pr(\epsilon_k \leq \epsilon_j + p_k - p_j \forall k \neq j \mid \epsilon_j) dP(\epsilon_j) && (19)\\
&= \int_{-\infty}^{\infty} \Pi_{k \neq j} F(\epsilon_j + p_k - p_j) f(\epsilon_j) d\epsilon_j && (20)\\
&= \int_{-\infty}^{\infty} F^{J-1}(\epsilon_j) f(\epsilon_j) d\epsilon_j && (21)\\
&= 1/J && (22)\\
\Rightarrow & && (23)\\
\frac{\partial s_j}{\partial p_j} &= -\int_{-\infty}^{\infty} \sum_{k \neq j} \left(f(\epsilon_j + p_k - p_j) \Pi_{l \neq k,j} F(\epsilon_j + p_k - p_j) \right) f(\epsilon_j) d\epsilon_j && (24)\\
&= -\int_{-\infty}^{\infty} (J-1) F^{J-2}(\epsilon_j) f(\epsilon_j) f(\epsilon_j) d\epsilon_j && (25)\\
\Rightarrow & && (26)\\
1/markup &= J \int_{-\infty}^{\infty} (J-1) F^{J-2}(\epsilon_j) f(\epsilon_j) f(\epsilon_j) d\epsilon_j && (27)
\end{aligned}
$$

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

$$
\begin{aligned}
1/markup &= E_{\epsilon_2^J}[Jf(\epsilon)] && (28)\\
&\geq Jk_J Pr_{\epsilon_2^J}[Jf(\epsilon) \geq Jk_J] && (29)\\
&= Jk_J Pr_{\epsilon_2^J}[f(\epsilon) \geq k_J] && (30)\\
&= Jk_J \int_{-\infty}^{\infty} \{f(\epsilon) \geq k_J\} (J-1) F^{J-2}(\epsilon) f(\epsilon) d\epsilon && (31)
\end{aligned}
$$

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

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

$f(\epsilon) \geq k_J$. Now consider the integral above:

$$Jk_J \int_{-\infty}^{\infty} \{f(\epsilon) \geq k_J\}(J-1)F^{J-2}(\epsilon)f(\epsilon)d\epsilon \geq Jk_J \int_{\underline{\epsilon}_M}^{\infty} \{M(1-F(\epsilon)) \geq k_J\}(J-1)F^{J-2}(\epsilon)f(\epsilon)d\epsilon \tag{32}$$

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

$$M(1-F(\epsilon)) \geq k_J \tag{33}$$

$$\Leftrightarrow F(\epsilon) \leq 1 - \frac{k_J}{M} \tag{34}$$

$$\Leftrightarrow \epsilon \leq F^{-1}(1 - \frac{k_J}{M}) \tag{35}$$

Plugging this back into the integral gives:

$$Jk_J \int_{-\infty}^{\infty} \{f(\epsilon) \geq k_J\}(J-1)F^{J-2}(\epsilon)f(\epsilon)d\epsilon \geq Jk_J \int_{\underline{\epsilon}_M}^{F^{-1}(1-\frac{k_J}{M})} (J-1)F^{J-2}(\epsilon)f(\epsilon)d\epsilon \tag{36}$$

$$= Jk_J \left[(1 - \frac{k_J}{M})^{J-1} - F^{J-1}(\underline{\epsilon}_M)\right] \tag{37}$$

$$= Jk_J \left(1 - \frac{k_J}{M}\right)^{J-1} - Jk_J\delta^{J-1} \tag{38}$$

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

7 References

Ackerberg, D. and M. Rysman (2001), "Unobserved Product Differentiation in Discrete Choice Models: Estimating Price Elasticities and Welfare Effects," Working Paper, Boston University.

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

Bajari, P., and C. L. Benkard (2003), "Demand Estimation with Heterogenous Consumers and Unobserved Product Characteristics: A Hedonic Approach," Stanford University Working Paper.

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

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

Domencich, T. A. and D. McFadden (1975), *Urban Travel Demand,* Amsterdam: North Holland Publishing Company.

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

Manski, C. (1977), "The Structure of Random Utility Models," *Theory and Decision*, Vol. 8, 229-254.

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