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

October 2001

Abstract

We derive some theoretical economic properties of standard discrete choice econometric models that we believe are undesirable if the models are to be used as structural models of demand. We show that many standard models have the following properties: as the number of products increases, the compensating variation for removing all of the inside goods tends to infinity, all firms in a Bertrand-Nash pricing game have markups that are bounded away from zero, and for each good there is always some consumer that is willing to pay an arbitrarily large sum for the good. These undesirable properties may lead to incorrect conclusions about many policies of interest, including calculation of price indexes, the benefits of new goods, and the welfare loss due to mergers. We demonstrate that these undesirable properties hold not only in the logit model, but also in all random utility models with the following three general 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. One approach to avoiding these undesirable properties is to weaken these three restrictions. However, we also show that random utility models are not in general non-parametrically identified from market shares or individual level choice data. Our findings support the use of alternative structural approaches that have better economic properties, such as those of Bajari and Benkard (2001) and Berry and Pakes (2000).

*We would like to 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. We also thank seminar participants at Carnegie Mellon, Northwestern, Stanford, UBC, UCLA, Wash. Univ. in St. Louis, Wisconsin, and Yale. 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.

While random utility models are potentially very flexible models of consumer behavior, as a practical matter, many restrictions are imposed in order to simplify the complicated integrals that must be calculated when evaluating the likelihood function for these models. In particular, it is common to make strong parametric and independence assumptions about the distribution of the idiosyncratic error term. In the logit model, the error term is assumed to be \(iid\) extreme value, and in the multinomial probit model, the error term is assumed to be normally distributed. Such functional form assumptions are needed in applications to make estimation of the model computationally feasible. In this paper, we investigate the structural economic implications of common assumptions on the error term in random utility models. We identify several functional form assumptions that are typically used in empirical applications of random utility models that lead to counter-intuitive economic implications.

Our interest in this topic was prompted by some recent papers in empirical I.O. which find that some variants of the logit model appear to overstate the welfare benefits to product variety. For example, Petrin (1998) 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, Ackerberg and Rysman (2001) in studying the demand for phone books find that the optimal number of phone books 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.

The main assumption driving these counter-intuitive results is that the conditional support of the error term is unbounded. In section 3, we demonstrate that, if the utility function
satisfies standard continuity and monotonicity assumptions, then this property leads to six implications that are a priori undesirable in many applications.

The first implication is that the demand for all products is positive for all sets of prices, implying 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, 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 we demonstrate in the paper, this can lead to counter-intuitive implications for substitution patterns.

Third, in GEV based models, if the number of products becomes infinite, then no product has a perfect substitute. That is, each individual would almost surely suffer a utility loss bounded away from zero if she is forced to consume a product other than her preferred alternative. This property demonstrates that, in commonly used random utility models, 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, if the number of products tends toward infinity, a Bertrand-Nash equilibrium does not tend to-wards the perfectly competitive outcome where all firms price at marginal cost. This demonstrates that random utility models always build in “excess” market power for the firms.

Fifth, as the number of products tends to-wards infinity, the ratio of the error term to the deterministic part of utility at the maximum product becomes one. That is, all of utility is explained by the random error term.

Sixth, as the number of products tends to infinity, 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 above.

One reaction to these results is that econometricians should strive to weaken the functional form assumptions of standard random utility models. However, in section 4, we demonstrate that without fairly stringent functional form restrictions on the distribution of random coefficients, the utility function, and the random error terms, random utility models are not identified. This result shows that some functional form assumptions, implicit or explicit, are necessary in order to identify the structural demand model.

Thus, our results support the use of structural models in which willingness to pay estimates do not depend so heavily on realizations of the error term in the random utility model. For example, Berry and Pakes (2000) and Bajari and Benkard (2001) develop discrete choice models with heterogeneity in the utility function, but which do not have a random error term.\footnote{Note that the models of both these papers do have an additively separable error term. However, the error term is treated as fixed, not random.} 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.

Our research builds on two past literatures in microeconomics and applied microeconomics. First, there is a literature that has documented very large willingness to pay in many discrete choice models, particularly the logit model. Examples include Petrin (1998) and Ackerberg and Rysman (2001). Second, there is a literature that has studied the theoretical properties of discrete choice models, for example, Caplin and Nalebuff (1991) and Anderson, DePalma and Thisse (1991).

\section{The Model}

In this section, we develop a fairly general semi-parametric discrete choice model. This model nests as special cases many commonly used random utility models such as the logit, nested
logit, GEV, multinomial probit, as well as random coefficients versions of these models, including 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 over products can be written as a function of individual characteristics (describing individual tastes), product characteristics, and an additively separable random error term:

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

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 \( \mathcal{X} \subseteq \mathbb{R}^K \) is a compact set. In addition, \( p_j \in \mathbb{R}^+ \) is the price of product \( j \), \( y_i \in \mathbb{R}^+ \) is the income of consumer \( i \), \( \beta_i \) is a vector of individual taste parameters with support \( \mathcal{B} \subseteq \mathbb{R}^B \), and \( \epsilon_{ij} \) 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.\(^2\) The function \( u(\cdot) \) is assumed to take a known parametric form that is constant across individuals. Note that for the purposes of this section we place no restrictions on the joint distribution of \( \beta, y, \) and \( \epsilon \).

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

\(^2\) The price of the composite commodity is normalized to one.
In the model, consumers are rational utility maximizers. Consumer $i$ chooses product $j$ if and only if $j$ maximizes utility,

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

(3)

The random error terms $\epsilon_{ij}$ in equations (1)-(3) is typically interpreted as arising from a lack of information available to the econometrician due to the fact that the econometrician imperfectly observes some factors which influence consumer demand. Manski (1977) suggests that there are four sources of uncertainty which justify the use of the random error term in the model (1)-(5):

1. **Unobservable 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 mis-measured by the economist.

4. **Functional misspecification.** The econometrician typically does not know the true functional form for $u_{ij}$.

Introducing the random error term is convenient because under commonly used assumptions it guarantees that all choices have positive probability for all consumers. This is not true in many models without random error terms.

Let $s_{i,j} \equiv P_i(j|\beta_i, y_i)$ denote the probability that consumer $i$ chooses product $j$ conditional on $\beta_i$ 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:
\[ P_i(j|\beta_i, y_i) = \int_{-\infty}^{\infty} \int_{-\infty}^{u_{i,j} - u_{i,0}} \cdots \int_{-\infty}^{u_{i,j} - u_{i,j-1}} \int_{-\infty}^{u_{i,j} - u_{i,j+1}} \cdots \int_{-\infty}^{u_{i,j} - u_{i,J}} f(\epsilon|\beta_i, y_i)d\epsilon_{i,-j}d\epsilon_{i,j} \] (4)

That is, the probability that consumer \( i \) chooses product \( j \) is the probability that the realization of \( \epsilon \) 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. The object of interest is \( F(\beta, y, \epsilon) \), which is typically estimated with some parametric and independence restrictions.\(^3\) We discuss these restrictions in detail in the next section.

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

\[ P(j) = \int P_i(j|\beta_i, y_i)dF(\beta_i, y_i) \] (5)

In equation (5) we integrate out over the population distribution of \( \beta \) 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.

**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. Moreover, it also allows for much more general specifications. For a set of necessary

\(^3\) Bajari and Benkard (2001) also show how \( F(\beta, y, \epsilon) \) can be estimated non-parametrically using micro data.
and sufficient conditions for a demand system to be rationalized by random utility models see Anderson, DePalma and Thisse (1992) and McFadden and Richter (1990).

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 $\epsilon_{ij}$ is iid, independent of $\beta_i$ 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 $\epsilon_{ij}$ is independent of $\beta_i$ and $y_i$ and normally distributed. In that case, simulation methods such as GHK and Gibbs sampling can be used to compute (4). The fixed coefficient logit and probit models use even stricter assumptions.

In this section we demonstrate that many of 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. We focus specifically on assumptions made about the joint distribution of the unobservables with income, $F(\beta, y, \epsilon)$. We find that many of these economic properties are aggravated when the number of products, $J$, changes or becomes large.

Note that we are not the first to raise some of these issues (see for example, Andersen, de Palma and Thisse (1992), Berry, Levinsohn, and Pakes (1995), Berry and Pakes (2000), Caplin and Nalebuff (1991), and Petrin (1998)). However, many of the properties listed here are new to the literature, and furthermore we have attempted to show all of our results formally in order to clarify exactly which assumptions are at fault in each case.
3.1 Assumptions

In this subsection we list three assumptions that we will maintain throughout this section of the paper. The first assumption we make is a commonly used assumption regarding the independence of the consumer taste coefficients and the error terms.

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

This assumption is made for convenience only and, 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 list conditions on the set of utility functions that we consider.

**Assumption U**

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

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

Assumption U(i) says that all individuals’ utility functions are continuous and that all 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 $\delta_M$ such that $\Pr(\epsilon_{ij} < M|\epsilon_{i,-j}) < \delta_M < 1$ for all $i, j$ pairs, all $\epsilon_{i,-j}$, and all $J \in \mathbb{Z}_+$.  

Assumption R amounts to assuming that the conditional error distributions have unbounded upper support and a continuous upper tail. 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.  

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 $\mathbb{R}^+$. Then in the model described above, for every product, demand is 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

\footnote{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 which has a finite price, thus placing an upper bound on the amount any consumer should be willing to pay for the good in question. 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 (2001) 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

---

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

6 For example, this would happen if the two firms had different costs, or if the two firms were multi-product firms with different product lines and one identical product.
terms. In many markets this property may be unreasonable.

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.

Suppose that \( u(x_j, y_i - p_j, \beta_i) \) has the form:

\[
\alpha \cdot y_i - \beta \cdot x_j - \alpha(y_i - p_j)
\]

Property two shows that if the market share of the outside good is bounded away from zero, as would typically be the case, then in the limit \( \beta \cdot x_j - \alpha(y_i - p_j) \) must tend toward negative infinity for all but finitely many products. This will clearly have pathological implications for many uses of the model since it implies either that one or more of the characteristics has negative marginal utility, or that \( \alpha \) is infinite.
If we are trying to describe choice behavior in a narrowly defined market, it likewise 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.

We now list one additional assumption regarding the error distribution:

**Assumption H** For each $j$, the limit as $\epsilon_j$ tends to the upper limit of its support of the hazard rate of $F_{\epsilon_j}(\cdot)$ is infinite, i.e., that $\lim_{\epsilon \to b} \frac{f_{\epsilon_j}(\epsilon)}{1 - F_{\epsilon_j}(\epsilon)} = \infty$, where $b$ is the upper end of the support of $F_{\epsilon_j}(\cdot)$ 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 $\epsilon_{ij}$ 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 $\epsilon_{ij}$ 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.\(^7\)

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 \(\epsilon_{ij}\), 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

\(^7\) Anderson, de Palma, and Thisse also show this for the standard logit model.
seems economically unintuitive. It seems more intuitive that the percentage of the variance in utility explained by observables should remain more or less constant for any given market as the number of products in the market changes.

6. **Compensating Variation** Suppose that assumptions I, R, and U hold. Then as the number of products becomes large the compensating variation for removing all of the inside goods almost surely tends to infinity for every individual.

Property 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:

\[ \epsilon_{ij} = \lambda_i' \eta_j \]  

where \( \lambda_i \) is individual \( i \)'s \( J \)-dimensional random vector of tastes for each product, and \( \eta_j \) is a vector of zeros with a one in the \( j \)th 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 counterintuitive 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 $I \times J$, where $I$ is the number of individuals and $J$ is the number of products. That is, the error vector consists of $I \times J$ 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 (1998)).

4 Identification

Since the results of the previous section demonstrate that the standard assumptions imposed in random utility models in order to ease estimation can lead to an economic model with undesirable properties, a first reaction may be to attempt to weaken the parametric restrictions in estimation. In this section, we demonstrate that it is not possible to estimate the model in full generality because the model (1)-(5) is not identified from market share data without making functional form assumptions about the primitives. Our proofs are constructive and our results can easily be extended to show that the model is not identified even if individual level choice data is observed.
Let \( P(j) \) be the market share of product \( j \) as defined in equation (5). We say that our model is identified if it is possible to uniquely recover the primitives of the model, \((u, F_{\beta, y}, F_{\epsilon})\) from observing the vector of market shares \( s \).

**Definition** We say that the primitives of our model are identified if \((u, F_{\beta, y}, F_{\epsilon}) \neq (u', F'_{\beta, y}, F'_{\epsilon})\) implies \( s \neq s' \).

We prove that the primitives of our model are not identified by showing that it is always possible to rationalize the observed market shares even if \( \epsilon_j \) is always zero. Suppose that preferences can be expressed in the following form:

\[
u(x_j, y_i - p_j, \beta_i) = -\sum_{k=1}^{K} (x_{j,k} - \beta_{i,k})^2 (7)\]

This is sometimes known as the “bliss point” model of preferences. That is, consumer \( i \)'s preferences are a \( K \) dimensional vector \( \beta_i = (\beta_{i,1}, ..., \beta_{i,K}) \) and the consumer seeks to minimize the norm of the distance between her preference vector and the product that she chooses. If we let \( \beta_i = x_j \) then obviously consumer \( i \)'s most preferred product will be \( j \). If \( \beta_i = x_j \) with probability \( s_j \), then obviously we can perfectly rationalize the observed market shares with our bliss point model. Moreover, suppose that the first characteristic is continuous and not equal for all of the \( j \) products in our data set. If \( \beta_{i,1} = x_{j,1} \) and \( \beta_{i,k} = 0 \) for \( k \neq 1 \) then product \( j \) is still utility maximizing for consumer \( i \). By setting \( \beta_{i,1} = x_{j,1} \) with probability \( s_j \) and \( \beta_{i,k} = 0 \) for \( k \neq 1 \) we are able to perfectly rationalize the observed market shares.

Alternatively, suppose that our model of preferences is:

\[
u(x_j, y_i - p_j, \beta_i) = \epsilon_{i,j} (8)\]

If we let \( \epsilon_{i,j} \) be \( e_j \), the unit vector with 1 in the \( j \)th position and zero elsewhere, then product \( j \) is utility maximizing for consumer \( i \). If we let \( \epsilon_{i,j} \) be \( e_j \) with probability \( s_j \), then we can perfectly rationalize the observed market shares.
It is worth noting that by allowing the distribution of $e_j$ to change over time, we can rationalize any time series variation in market shares using model (8). Similarly, if we allow the distribution of $\beta_i$ to change over time we can perfectly rationalize any market shares using model (7).

These results suggest that it is not possible to identify individual level preferences from aggregate level data without further restrictions on the model primitives $(u, F_{\beta, y}, F_{\epsilon})$.

**Theorem 1.** The primitives of our discrete choice model, $(u, F_{\beta, y}, F_{\epsilon})$ are not identified from observing market shares.

The bliss point specification is not the only specification that can be used to rationalize the data. In our companion paper, Bajari and Benkard (2001), we demonstrate that so long as the taste vector $\beta$ is at least $K$-dimensional and certain regularity conditions hold, a large number of functional forms can be used to rationalize the observed data without including the random error term $\epsilon_{i,j}$.

In our view, these results should not be particularly surprising because standard data sets typically contain no more than one choice observation per individual. The information contained such data sets is not enough to fully identify individual preferences without functional form or homogeneity restrictions.

Furthermore, in economic applications it is often necessary to make functional form assumptions about the primitives in order to achieve identification. For example, Rust (1991) notes that dynamic discrete choice models are not identified without making assumptions about the period return function. Similarly, the results of who? (197?) demonstrate that it is not possible to recover consumer tastes in general equilibrium models just from knowledge of the aggregate demand function.

We interpret our results as indicating that it is critical to think carefully about which as-
sumptions are economically reasonable when estimating demand in differentiated product markets. It is not necessarily possible to recover the full structural demand system by simply estimating a discrete choice model and “letting the data speak for itself”.

5 Alternative Models

In section 3, we demonstrated that many commonly used random utility models 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 as in Ackerberg and Rysman (2001).\footnote{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.} In this section, we demonstrate that the pure hedonic models studied in Berry and Pakes (2000) and Bajari and Benkard (2001) are another approach to avoiding some of the unappealing properties of commonly used discrete choice models.

Hedonic models of demand for differentiated products have been used extensively in industrial organization. Examples include models of horizontal product differentiation such as Hotelling (1929), Gorman (1980) and Lancaster (1966), models of vertical product differentiation, such as Shaked and Sutton (1987) and Bresnahan (1987), as well as Rosen’s (1974) model which is the basis of the two-stage hedonic approach.

In the pure hedonic model, commodities are fully described by a collection of a finite number of attributes, which we denote as $x_j$ as above. 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 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:
\[
\max_{(j,c)} u_i(x_j, \xi_j, c) \\
\text{s.t.} \quad p_j + c \leq y_i
\] (9) (10)

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:

\[
\begin{align*}
  u_{ij} &= \beta_{1,i}x_{j,1} + \beta_{2,i}x_{j,2} + \beta_{3,i}x_{j,3} + (y_i - p_j) 
\end{align*}
\] (11)

Note that in (11) we assume 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 (11), each consumer \(i\) has a unique taste coefficient for each characteristic. In Bajari and Benkard (2001), we demonstrate 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 (2001) also show that the taste coefficients, \(\beta_i\) will typically be just identified if there are as many taste coefficients as product characteristics.

The hedonic model of Bajari and Benkard (2001) is similar to many commonly used random utility models except for the following two features:

1. There is no random idiosyncratic taste shock \(\varepsilon_{ij}\).
2. No parametric or independence restrictions are imposed on the joint distribution of \(\beta_i\).
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:

$$x_{j',k} > x_{j,k} \text{ for } k = 1, ..., K$$  \hspace{1cm} (12)

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 if that case.

6 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.

These undesirable properties may lead to incorrect conclusions in applications such as price indexes, the welfare gains from new products, or in analyzing the effects of mergers in a differentiated products market. They may also explain the findings of Petrin (1998) and Ackerberg and Rysman (2001), who have found that the logit model tends to overstate the benefits of product variety.

In this paper, we have demonstrated that these properties will hold not only in the logit model, but in any random utility model with the following two 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 random utility models, properties (1)-(3) are maintained in most applications we are aware of.

One approach to avoiding these undesirable properties is to weaken the parametric restrictions (1)-(3). However, the random utility model in its general form is not non-parametrically identified from market shares or individual level choice data. This follows from the fact that it is possible to rationalize any data set with a variety of alternative models.

Our results support the use of alternative 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 (2001), do not have the undesirable properties derived in section 3.
7 Appendix

7.1 Proofs for Section 2

7.1.1 Proof of Property 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 \(\beta\). 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

\[
\epsilon_{ij} > \max_{k \neq j} \{ u(x_k, \bar{y}_i - p_k, \bar{\beta}_i) \} - u(x_j, \bar{y}_i - p_j, \bar{\beta}_i) \equiv \bar{u}_k - \bar{u}_j.
\] (13)

By R, the probability corresponding to (13) is strictly positive.

Let

\[
A_j = \{(y, \beta, \epsilon) \in \mathbb{R}_+ \times B \times \mathbb{R}_+^{J+1} | \ u_{ij} \geq u_{ik} \ \forall k \in 0..J \}
\] (14)

\(A_j\) 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, \(M\).

\[
q_j(x, p; \theta) = M \int_{\mathbb{R}_+^{J+1}} \int_B \int_{\mathbb{R}_+} A_j \ dF(\beta, y) dF(\epsilon)
\] (15)

Thus, using the same point above,

\[
q_j(x_j, p_j, x_{-j}, p_{-j}) > M \cdot Prob[\epsilon_{ij} > \bar{u}_k - \bar{u}_j|\bar{\beta}, \bar{y}_i] f(\bar{\beta}, \bar{y}_i) > 0
\] (16)

Since we chose the vector of prices arbitrarily, demand is positive for every good for every price vector. \(\square\)
7.1.2 Proof of Properties 2 and 5

Lemma 2. Assumption R implies that \( \lim_{J \to \infty} \max_{j \in 1..J} \epsilon_i = \infty \) a.s.

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

\[
Pr(A_n) = Pr(\epsilon_{i0} < M)Pr(\epsilon_{i1} < M|\epsilon_{i0} < M) \ast ... \ast Pr(\epsilon_{in} < M|\epsilon_{i-n} < M)
\] (17)

By assumption R, there exists a \( \delta_M \) such that each term in the above expression is less than \( \delta_M \). 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.

7.1.3 Property 6

Proof of property 6. By the previous result,

\[
\lim_{J \to \infty} \max_{j \in 1..J} \epsilon_{ij} = \infty
\] (18)

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

\[
u(0, y_i + CV, \beta_i) = \max_{j \in 1..J} \left\{ u(x_j, y_i - p_j, \beta_i) + \epsilon_{ij} \right\} - \epsilon_{i0}
\] (19)

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.
7.1.4 Properties 3 and 4

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

We show two proofs: 1) \( \lim_{J \to \infty} E[\epsilon_{J1} - \epsilon_{J2}] = 0 \), where \( \epsilon_{J1} \) is the highest of \( J \) draws on \( \epsilon \) and \( \epsilon_{J2} \) 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 \( E_{\epsilon_2} [\lim_{J \to \infty} E(\epsilon_{J1} - \epsilon_{J2} \mid \epsilon_{J2})] \). Now, note that we have shown above that \( \lim_{J \to \infty} \epsilon_{J2} = \infty \) 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 \( y \mid x \) goes to infinity as \( x \) becomes large. But, the conditional distribution of \( \epsilon_{J1} \mid \epsilon_{J2} \) is proportional to the distribution of \( \epsilon \). Thus \( \lim_{J \to \infty} E[\epsilon_{J1} - \epsilon_{J2} \mid \epsilon_{J2}] = 0 \) 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 $s_j = \frac{s_j}{p_j}$. We now consider $\frac{\partial s_j}{\partial p_j}$:

\[ s_j = Pr(\epsilon_k \leq \epsilon_j + p_k - p_j \forall k \neq j) \]

\[ = \int_{-\infty}^{\epsilon_j} Pr(\epsilon_k \leq \epsilon_j + p_k - p_j \forall k \neq j | \epsilon_j) dP(\epsilon_j) \]

\[ = \int_{-\infty}^{\epsilon_j} \prod_{k \neq j} F(\epsilon_j + p_k - p_j) f(\epsilon_j) d\epsilon_j \]

\[ = \int_{-\infty}^{\epsilon_j} F^{J-1}(\epsilon_j) f(\epsilon_j) d\epsilon_j \]

\[ = 1/J \]

\[ \Rightarrow \]

\[ \frac{\partial s_j}{\partial p_j} = -\int_{-\infty}^{\epsilon_j} \sum_{k \neq j} (f(\epsilon_j + p_k - p_j) \prod_{l \neq k,j} F(\epsilon_j + p_k - p_l)) f(\epsilon_j) d\epsilon_j \]

\[ = -\int_{-\infty}^{\epsilon_j} (J - 1) F^{J-2}(\epsilon_j) f(\epsilon_j) f(\epsilon_j) d\epsilon_j \]

\[ \Rightarrow \]

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

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 $\epsilon^J$ so that the whole expression can be written as $E_{\epsilon^J} [J f(\epsilon)]$.

By Markov’s inequality, we have:

\[ 1/\text{markup} = E_{\epsilon^J} [J f(\epsilon)] \geq J k_J P_{\epsilon^J} [J f(\epsilon) \geq J k_J] \]

\[ = J k_J P_{\epsilon^J} [f(\epsilon) \geq k_J] \]

\[ = J k_J \int_{-\infty}^{\epsilon^J} \{f(\epsilon) \geq k_J\} (J - 1) F^{J-2}(\epsilon) f(\epsilon) d\epsilon \]

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

\[
Jk_J \int_{-\infty}^{\infty} \{ f(\epsilon) \geq k_J \} (J - 1) F^{J-2}(\epsilon) f(\epsilon) d\epsilon \geq Jk_J \int_{\xi_M}^{\infty} \{ M(1 - F(\epsilon)) \geq k_J \} (J - 1) F^{J-2}(\epsilon) f(\epsilon) d\epsilon
\]

(34)

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

\[
M(1 - F(\epsilon)) \geq k_J
\]

(35)

\[
\Leftrightarrow F(\epsilon) \leq 1 - \frac{k_J}{M}
\]

(36)

\[
\Leftrightarrow \epsilon \leq F^{-1}(1 - \frac{k_J}{M})
\]

(37)

Plugging this back into the integral gives:

\[
Jk_J \int_{-\infty}^{\infty} \{ f(\epsilon) \geq k_J \} (J - 1) F^{J-2}(\epsilon) f(\epsilon) d\epsilon \geq Jk_J \int_{\xi_M}^{F^{-1}(1 - \frac{k_J}{M})} (J - 1) F^{J-2}(\epsilon) f(\epsilon) d\epsilon
\]

(38)

\[
= Jk_J \left[ (1 - \frac{k_J}{M})^{J-1} - F^{J-1}(\xi_M) \right]
\]

(39)

\[
= Jk_J \left( 1 - \frac{k_J}{M} \right)^{J-1} - Jk_J \delta^{J-1}
\]

(40)

where \( \delta = F(\xi_M) < 1 \). 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. \( \square \)
8 References


