Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers*

Steven T. Berry
Yale University
Department of Economics
Cowles Foundation
and NBER

Philip A. Haile
Yale University
Department of Economics
Cowles Foundation
and NBER

June 10, 2008
Preliminary and Incomplete

Abstract

We consider identification of nonparametric random utility models of multinomial choice using observation of consumer choices. Our model of preferences nests random coefficients discrete choice models widely used in practice with parametric functional form and distributional assumptions. However, our model is nonparametric and distribution free. It incorporates choice-specific unobservables and endogenous choice characteristics, both of which are essential to modeling demand in most settings. It also permits unknown heteroskedasticity and correlated taste shocks. Under standard orthogonality, "large support," and instrumental variables assumptions, we show identifiability of choice-specific unobservables and the joint distribution of preferences conditional on any vector of observed and unobserved characteristics. We demonstrate robustness of these results to relaxation of the large support condition and show that when this condition is replaced with a much weaker "common choice probability" condition, the demand structure is still identified. We show that our key maintained hypotheses are testable.

*We had helpful early conversations on this topic with Rosa Matzkin and Yuichi Kitamura. We also thank Sunyoung Park and seminar participants at Chicago, Harvard, Northwestern, M.I.T., and Yale for helpful comments.

1 Introduction

We consider identification of nonparametric random utility models of multinomial choice using observation of consumer choices, i.e., "micro data."[1] Our model of preferences nests random coefficients discrete choice models widely used in practice with parametric functional form and distributional assumptions. However, our model is nonparametric and distribution free. It incorporates choice-specific unobservables and endogenous choice characteristics, both of which are essential to modeling demand in most settings. It also permits unknown heteroskedasticity and correlated taste shocks. Under standard orthogonality, "large support," and instrumental variables assumptions, we show identifiability of choice-unobservables and of the joint distribution of preferences conditional on any vector of observed and unobserved characteristics. We demonstrate robustness of these results to relaxation of the large support condition and show that when this condition is replaced with a much weaker "common choice probability" condition (defined below), the demand structure is still identified. We also show that our key maintained hypotheses are testable.

Motivating our work is the extensive use of discrete choice models of demand for differentiated products in a wide range of applied fields of economics and related disciplines. Important examples include transportation economics (e.g., Domenich and McFadden (1975)), industrial organization (e.g., Berry, Levinsohn, and Pakes (2004)), international trade (e.g., Goldberg (1995)), marketing (e.g., Guadagni and Little (1983)), urban economics (e.g., Bayer, Ferreira, and McMillan (2007)), education (e.g., Hastings, Staiger, and Kane (2007)), migration (e.g., Kennan and Walker (2006)), political science (e.g., Rivers (1988)), and health economics (e.g., Ho (2007)). We focus in particular on discrete choice random utility models with unobserved characteristics in the spirit of Berry (1994), Berry, Levinsohn, and Pakes (1995) and a large related literature. Although this class of models has been applied to research in many areas, the sources of identification of these models have not been fully understood. Without such an understanding it is difficult to know what qualifications are necessary when interpreting estimates or policy conclusions.

[1] We consider identification using market level data in our companion paper, Berry and Haile (2008).

Our analysis demonstrates that with sufficiently rich data, random utility multinomial choice models featuring unobserved characteristics are identified without the parametric assumptions used in practice – typically, linear utility with independent taste shocks (entering additively and/or multiplicatively) drawn from parametrically specified distributions. Indeed, we provide positive identification results for more general models of preferences than those considered previously (to our knowledge) in the econometrics or applied literatures.

Our results may therefore lead to greater confidence in estimates and policy conclusions obtained using estimates of discrete choice demand models. In particular, parametric specifications used in practice can often be viewed as parsimonious approximations in finite samples rather than essential maintained assumptions. We view this as our primary message. However, our results also suggest that with large samples even richer specifications (parametric or nonparametric) of preferences might be considered in empirical work, and our identification proofs may suggest estimation approaches.

The identifiability of random utility discrete choice models is not a new question, and our results build on two well-known ideas (we relate our results more precisely to the prior literature in section 7). The first is that of a "special regressor"— an observable (or vector of observables) with large support (e.g., Manski (1985), Matzkin (1992), Lewbel (2000)). Following standard arguments, sufficient variation in such observables enables one to "trace out" the distribution of the random component of utilities for all choices within a given choice set. The second idea is the use of variation in choice characteristics, within and across choice sets, to decompose variation in the distribution of utilities into the contributions of observed and unobserved factors (e.g., Berry (1994), Berry, Levinsohn, and Pakes (1995), Berry, Levinsohn, and Pakes (2004)). Combining and extending these ideas enables us to obtain positive results for a less restrictive nonparametric model than those considered previously. In particular, the use of within-market variation in consumer attributes allows us to trace out a very flexible joint distribution while unobservable product characteristics are held fixed. Cross-market (and cross-product) variation in choice characteristics then allows identification of choice-specific unobservables.[2]

[2] In a fully parametric context, a similar intuition is suggested by Berry, Levinsohn, and Pakes (2004).

Although the generality of our model is a strength, it also brings some limitations. One is a constraint on the types of out-of-sample counterfactuals that can be identified. This is a constraint inherent to nonparametric models: functional form assumptions will often be required if one is to extrapolate outside the support of the data generating process. A second limitation is that our model lacks sufficient structure to permit full characterizations of welfare. In particular, we do not assume a structure that would enable one to track a given consumer's position in the distribution of indirect utilities across environments. Thus, although we will be able to identify changes in utilitarian social welfare (in aggregate or across subpopulations defined by observables), characterization of other welfare measures (e.g., Pareto improvements) requires additional structure on preferences. We provide additional discussion of these limitations below.

In the following section we set up the model and define the structural features of interest. In section 3 we demonstrate one of our main lines of argument for a simple case: binary choice with exogenous characteristics. Section 4 then addresses multinomial choice with endogeneity, considering two alternative instrumental variables conditions. We then move to discussion of several important extensions. In section 5 we discuss the case in which the large support condition fails. Section 6 then presents testable restrictions of key maintained hypotheses. Having presented our results, we are then able to place our contribution within the context of the large literature on identification of multinomial choice models. After doing this in section 7, we conclude in section 8.

2 Model

2.1 Setup

Consistent with the motivation from demand estimation, we describe the model as one in which each consumer i in each market t chooses from a set $\mathcal{J}_t$ of available products. We will use the terms "product," "good," and "choice" interchangeably to refer to elements of the choice set. The term "market" here is synonymous with the choice set. In particular, consumers facing the same choice set can be viewed as being in the same market. In practice, markets will typically be defined geographically and/or temporally. Variation in the choice set will of course be essential to identification, and our explicit reference to markets provides a way to discuss this clearly.

In applications to demand it is important to model consumers as having the option to purchase none of the products considered (see, e.g., Bresnahan (1981), Anderson, DePalma, and Thisse (1992), Berry (1994) and Berry, Levinsohn, and Pakes (1995)). We represent this by choice $j = 0$ and assume $0 \in \mathcal{J}_t \; \forall t$. Choice 0 is often referred to as the "outside good." We denote the number of "inside goods" by $J_t = |\mathcal{J}_t| - 1$.[3] Each inside good j has observable (to us) characteristics x_{jt}, which may include price. Prices, of course, will generally be correlated with product-specific unobservables. Unobserved choice characteristics are characterized by an index ξ_{jt}, which may also vary across markets. We will assume that ξ_{jt} has an atomless marginal distribution in the population.

Each consumer i in market t is associated with a vector of observables z_{ijt}. The j subscript on z_{ijt} allows the possibility that some characteristics are consumer-choice specific—e.g., interactions between consumer demographics and product characteristics (say, family size and automobile size) or other consumer-specific choice characteristics (say, driving distance to retailer j from consumer i's home). For most of our results we will require at least one such measure for each $j \geq 1$ (we consider the case without micro data in Berry and Haile (2008)). We let $\mathbf{x}_t = (x_{1t}, \ldots, x_{J_t t})$ and $\mathbf{z}_{it} = (z_{i1t}, \ldots, z_{iJ_t t})$

We consider a random utility model. Consumers face no uncertainty themselves, but from the perspective of an outsider the preferences of any individual are viewed as random (e.g., Luce (1959), Block and Marschak (1960), McFadden (1974), Manski (1977)), with the usual interpretation that this reflects unobserved consumer-specific tastes for products and/or characteristics. Let $(\Omega, \mathcal{F}, \mathbb{P})$ denote a probability space. Each consumer i's preferences are assumed to be described by a random utility function

$$u(\cdot, \cdot, \cdot, \omega_{it}) : \mathbb{R}^{K_x} \times \mathbb{R} \times \mathbb{R}^{K_z} \to \mathbb{R}$$

[3] In applications with no "outside choice" our approach can be adapted by normalizing preferences relative to those for a given choice. The same adjustment applies when characteristics of the outside good vary across markets in observable ways.

where $\omega_{it} \in \Omega$, and u is a measurable function. Given a choice set $\mathcal{J}$ and $\left\{\left(x_{jt}, \xi_{jt}, z_{ijt}\right)\right\}_{j\in\mathcal{J}}$, consumer i's preferences are then determined by the conditional indirect utilities

$$v_{ijt} = u\left(x_{jt}, \xi_{jt}, z_{ijt}, \omega_{it}\right) \quad \forall i, j \tag{1}$$

Implicit in this formulation is a standard restriction that, conditional on the values of $\left(x_{jt}, \xi_{jt}, z_{ijt}\right)$, the random variation in the conditional indirect utilities is i.i.d. across individuals and markets. We mark this restriction explicitly with the following.

Assumption 1 *The measure $\mathbb{P}$ on Ω does not vary with i, t, $\mathcal{J}_t$, or $\left\{\left(x_{jt}, \xi_{jt}, z_{ijt}\right)\right\}_{j\in\mathcal{J}_t}$.*

The invariance of $\mathbb{P}$ to i describes the sampling structure: conditional on $\left\{x_{jt}, \xi_{jt}, z_{ijt}\right\}_{j\in\mathcal{J}_t}$, unobservable variation in the preferences of different individuals reflects independent draws of the elementary outcome ω_{it} from Ω. This does not rule out within-market correlation in consumers' preferences conditional on the observables$\{x_{jt}, z_{ijt}\}_{j\in\mathcal{J}_t}$, since ξ_{jt} can be interpreted as an unobserved market-level taste for good j. The invariance to $\mathcal{J}_t$ and $\left\{x_{jt}, \xi_{jt}, z_{ijt}\right\}_{j\in\mathcal{J}_t}$ (and thus to t) reflects the standard view of preferences as stable, rather than varying with the choice set. In particular, the realization of ω_{it} determines the utility function $u\left(\cdot,\cdot,\cdot,\omega_{it}\right)$ of a consumer, while the values of $\left\{x_{jt}, \xi_{jt}, z_{ijt}\right\}_{j\in\mathcal{J}_t}$ determined the relevant sets of arguments of this function. Note that this structure allows arbitrary heterogeneity in the stochastic component of utilities across consumers with different z_{ijt}.

As the following example illustrates, Assumption 1 does *not* require homoskedasticity or the common assumption that taste shocks for a given individual are mutually independent.[4]

Example 1 *A special case of our general framework is the linear random coefficients model*

$$u\left(x_{jt}, \xi_{jt}, z_{ijt}, \omega_{it}\right) = x_{jt}\beta_{it} + z_{ijt}\gamma + \xi_{jt} + \epsilon_{ijt} \tag{2}$$

[4] This example allows heterskedasticity in random utilities through to the linear random coefficients. In general our specification allows arbitrary heterskedasticity in v_{ijt} under Assumption 1.

where, for example, $\beta_{it} = (\beta^{(1)}_{it}(\omega_{it}), \ldots, \beta^{(k)}_{it}(\omega_{it}))$ *is a vector of random coefficients and each* $\epsilon_{ijt}(\omega_{it})$ *is a consumer-choice specific taste shock whose distribution varies with choice-specific observables. With this specification, Assumption 1 allows an arbitrary joint distribution of* $\left(\beta^{(1)}_{it}, \ldots, \beta^{(k)}_{it}, \epsilon_{i1t}, \ldots, \epsilon_{iJ_tt}\right)$ *but requires that this joint distribution be the same for all* i*,* t*, and* $\left\{(x_{jt}, \xi_{jt}, z_{ijt})\right\}_{j=1\ldots J}$.[5] *As an alternative, we could specify* $\beta_{it} = (\beta^{(1)}_{it}(z_{it}, \omega_{it}), \ldots, \beta^{(k)}_{it}(z_{it}, \omega_{it}))$ *and* $\epsilon_{ijt}(x_{jt}, \omega_{it})$*, where* z_{it} *is a vector of individual characteristics that do not vary across* j*. Now, for example, Assumption 1 requires that the joint distribution of* $(\epsilon_{i1t}, \ldots, \epsilon_{iJ_tt})$ *be the same for choice sets with identical observable characteristics.*

Each consumer i maximizes her utility by choosing good j whenever

$$u\left(x_{jt}, \xi_{jt}, z_{ijt}, \omega_{it}\right) > u\left(x_{kt}, \xi_{kt}, z_{ikt}, \omega_{it}\right) \qquad \forall k \in \mathcal{J}_t - \{j\}. \tag{3}$$

Denote consumer i's choice by

$$y_{it} = \arg\max_{j \in \mathcal{J}_t} \ u\left(x_{jt}, \xi_{jt}, z_{ijt}, \omega_{it}\right).$$

Let $z_{ijt} = \left(z^{(1)}_{ijt}, z^{(2)}_{ijt}\right)$, with $z^{(1)}_{ijt} \in \mathbb{R}$. Let $\mathbf{z}^{(1)}_{it}$ denote the vector $\left(z^{(1)}_{i1t}, \ldots, z^{(1)}_{iJ_tt}\right)'$ and $\mathbf{z}^{(2)}_{it}$ the matrix $\left(z^{(2)}_{i1t}, \ldots, z^{(2)}_{iJ_tt}\right)'$. We will require that for each possible $\mathbf{z}^{(2)}_{it}$, there exist a representation of preferences with the form

$$\tilde{u}_{ijt} = \phi_{it} z^{(1)}_{ijt} + \tilde{\mu}\left(x_{jt}, \xi_{jt}, z^{(2)}_{ijt}, \omega_{it}\right) \qquad \forall i, j = 1, \ldots, \mathcal{J}_t \tag{4}$$

for some function $\tilde{\mu}$ that is strictly increasing and continuous in ξ_{jt}, and with the random coefficient $\phi_{it} = \phi(\omega_{it})$ strictly positive with with probability one.[6] Here we have imposed two restrictions: (i) additive separability in a "vertical" component, $z^{(1)}_{ijt}$, of z_{ijt}, (ii) monotonicity in

[5] This structure permits variation in J_t across markets. The realization of ω_{it} should be thought of as generating values of $\epsilon_{ijt} = \epsilon_j(\omega_{it})$ for all possible choices j, not just those in the current choice set. Thus, preferences exist even over products not available. Note that here the joint distribution of $\{\epsilon_{ijt}\}_{j \in \mathcal{K}}$ will be the same regardless of whether $\mathcal{K} = \mathcal{J}_t$ or $\mathcal{K} \subset \mathcal{J}_t$. Thus, a consumer's preference between two products j and k does not depend on the other products in the the choice set.

[6] If $\phi_{it} < 0$ w.p. 1, we replace $z^{(1)}_{ijt}$ with $-z^{(1)}_{ijt}$. As long as $|\phi_{it}| > 0$ w.p. 1, identification of the sign of ϕ_{it} is straightforward under the assumptions below.

ξ_{jt}. We show in section 6 that both restrictions have testable implications.

We rely on the separability restriction to provide a mapping between units of (latent) utility and units of (observable) choice probabilities.[7] Because unobservables have no natural order, monotonicity in ξ_{jt} would be without loss of generality if consumers had homogeneous tastes for characteristics, as in standard multinomial logit, nested logit, and multinomial probit models. With heterogeneous tastes for choice characteristics, monotonicity imposes a restriction that ξ_{jt} be a "vertical" rather than "horizontal" choice characteristic. Thus, all consumers agree that (all else equal) larger values of ξ_{jt} are preferred. Of course, our specification does allow heterogeneity in tastes for ξ_{jt}, just as this is permitted for the vertical characteristic $z_{ijt}^{(1)}$. Furthermore, we allow a different representation (4) for each value of $\mathbf{z}_{it}^{(2)}$.[8]

We need to make several normalizations in order to obtain a unique representation of preferences from (4). First, because unobservables have no natural units we may normalize the location and scale of ξ_{jt} and assume without loss that it has a uniform marginal distribution on $(0,1)$. We must also normalize the location and scale of utilities. Without loss, we normalize the scale of consumer i's utility using his marginal utility from $z_{ijt}^{(1)}$, yielding the representation

$$u_{ijt} = z_{ijt}^{(1)} + \frac{\tilde{\mu}\left(x_{jt}, \xi_{jt}, z_{ijt}^{(2)}, \omega_{it}\right)}{\phi_{it}} \qquad \forall i, j = 1, \ldots, J^t.$$

Letting

$$\mu\left(x_{jt}, \xi_{jt}, z_{ijt}^{(2)}, \omega_{it}\right) = \frac{\tilde{\mu}\left(x_{jt}, \xi_{jt}, , z_{ijt}^{(2)}, \omega_{it}\right)}{\phi_{it}}$$

[7] We can extend this to allow $z_{ijt}^{(1)}$ to be an index. For example if $z_{ijt}^{(1)} = c_{ijt}\eta$, the parameter vector η can be identified up to scale directly from the observed choice probabilities as long as $\xi_{jt} \perp\!\!\!\perp c_{ijt}$.

[8] Athey and Imbens (2007) point out that the assumption of a scalar vertical unobservable ξ_{jt} can lead to testable restrictions in some models. In our model, if there were no variation across j in $z_{ijt}^{(1)}$ holding consumer characteristics fixed, consumers with the same $\mathbf{z}_{it}^{(2)}$ but different $\mathbf{z}_{it}^{(1)}$ must rank (probabalistically) any products with identical observable characteristics the same way, as they point out. Their observation does not apply to our model in general (e.g., conditional indirect utilities of the form $v_{ijt} = \xi_{jt} + z_{ijt}^{(1)}\beta_{it}$ are permitted by our model and do not lead to their their testable restriction), but we show below that there is a related testable restriction for our more general model.

this gives the representation of preferences we will work with below:

$$u_{ijt} = z_{ijt}^{(1)} + \mu\left(x_{jt}, \xi_{jt}, z_{ijt}^{(2)}, \omega_{it}\right) \qquad \forall i, j = 1, \ldots, J^t. \tag{5}$$

To normalize the location we set $u_{i0t} = 0\ \forall i, t$. Treating the utility from the outside good as non-stochastic is without loss of generality here, since choices in (3) are determined by differences in utilities and we have not restricted correlation in the random components of utility across choices.

Our model nests random utility models considered in applied work across a wide range of fields, including the following examples.

Example 2 *Consider the model of preferences for automobiles in Berry, Levinsohn, and Pakes (2004):*

$$\begin{aligned} u_{ijt} &= x_{jt}\beta_{it} + \xi_{jt} + \epsilon_{ijt} \\ \beta_{it}^k &= \beta_1^k + \beta_2^{k0}\nu_{it}^k + \sum_r z_{it}^r \beta_3^{kr} \quad k = 1, \ldots, K \end{aligned}$$

where $x_{jt} \in \mathbb{R}^k$ are auto characteristics, z_{it}^r are consumer characteristics, ϵ_{ijt} is assumed distributed type 1 extreme value, each ν_{it}^k is a standard normal deviate, and all stochastic components are i.i.d. Here β_1^k, β_2^{k0}, and β_3^{kr} are all parameters of our function μ in (5).

Example 3 *Consider the model of hospital demand in Capps, Dranove, and Satterthwaite (2003), where consumer i's utility from using hospital j depends on hospital characteristics x_{jt}, patient characteristics z_{it}, interactions between these, and patient i's distance to hospital j, denoted z_{ijt}. In particular,*

$$u_{ijt} = \alpha x_{jt} + \beta z_{it} + x_{jt}\Gamma z_{it} + \gamma z_{ijt} + \epsilon_{ijt}$$

with ϵ_{ijt} distributed type I extreme value.

Example 4 *Rivers (1988) considered the following model of voter preferences*

$$u_{ijt} = \beta_{1i}\left(z_{it}^{(1)} - x_{jt}^{(1)}\right)^2 + \beta_{2i}\left(z_{it}^{(2)} - x_{jt}^{(2)}\right)^2 + \epsilon_{ijt}$$

where $z_{it}^{(1)}$ *and* $x_{jt}^{(1)}$ *are, respectively, measures of voter i's and candidate j's political positions,* $z_{it}^{(2)}$ *and* $x_{jt}^{(2)}$ *are measures of party affiliation. Here the terms* $\left(z_{it}^{(1)} - x_{jt}^{(1)}\right)^2$ *and* $\left(z_{it}^{(2)} - x_{jt}^{(2)}\right)$ *form the consumer-choice specific observables we call* z_{ijt}.

2.2 Observables and Structural Features of Interest

When we discuss the case of endogenous choice characteristics we will require excluded instruments, which we denote by $\tilde{w}_{jt}$.[9] The observables consist of $(y_{it}, \{x_{jt}, \tilde{w}_{jt}, z_{ijt}\}_{j\in\mathcal{J}_t})_{i,t}$. To discuss identification, we treat their joint distribution as known. Loosely speaking, we consider the case of a large number of markets, each with a large number of consumers.

The observables directly reveal the conditional choice probabilities

$$p_{ijt} = p_j\left(\mathcal{J}_t, \{x_{jt}, \tilde{w}_{jt}, z_{ijt}\}_{j\in\mathcal{J}_t}\right) = \Pr\left(y_{it} = j | \mathcal{J}_t, \{x_{kt}, \tilde{w}_{kt}, z_{ikt}\}_{k\in\mathcal{J}_t}\right). \tag{6}$$

Although these alone reveal some important features of the model (e.g., average marginal rates of substitution between exogenous characteristics), they are not adequate for most purposes motivating demand estimation—for example, calculation of (own and cross-price) elasticities of demand. This is merely the standard observation that equilibrium prices and quantities do not identify demand.

Our first objective is to derive sufficient conditions for identification of the choice-specific unobservables and the distribution of preferences over choices in sets $\mathcal{J}_t$, conditional on the characteristics $\left\{x_{jt}, z_{ijt}, \xi_{jt}\right\}_{j\in\mathcal{J}_t}$. In particular, we will show identification of the joint distribution of $\{u_{i1t}\}_{j\in\mathcal{J}_t}$ conditional on any $\left(\mathcal{J}_t, \left\{x_{jt}, z_{ijt}, \xi_{jt}\right\}_{j\in\mathcal{J}_t}\right)$ in their support. These conditional distributions fully characterize the primitives of this model.

[9] Depending on the environment, instruments might include cost shifters excludable from the utility function, prices in other markets (e.g., Hausman (1996), Nevo (2001)), and/or characteristics of competing products (e.g., Berry, Levinsohn, and Pakes (1995)). Because the arguments are standard, we will not discuss assumptions necessary to justify the exlusion restrictions, which we will assume directly.

We will also consider a type of partial identification. For many economic questions motivating demand estimation, the joint distribution of utilities is not actually required. For example, to discuss cross-price elasticities, equilibrium pricing or market shares under counterfactual ownership or cost structures, knowledge of the demand structure alone is adequate. Identification of demand naturally requires less from the model and/or data than identification of the underlying distribution of preferences. In the multinomial choice setting, demand is fully characterized by the *structural choice probabilities*

$$\rho_j\left(\mathcal{J}_t, \{x_{jt}, \xi_{jt}, z_{ijt}\}_{j\in\mathcal{J}_t}\right) = \Pr\left(y_{it} = j | \mathcal{J}_t, \{x_{jt}, \xi_{jt}, z_{ijt}\}_{j\in\mathcal{J}_t}\right). \quad (7)$$

These conditional probabilities fully characterize demand; however, they are not directly observable from (6) because of the unobservables ξ_{jt}, which are typically correlated with at least some elements of x_{jt} (e.g., price). Some of our results below address identification of these choice probabilities, i.e., of the demand structure.

2.3 Some Limitations of the Model

The generality of our model of preferences comes with some costs. One is that we know before starting that some out-of-sample counterfactuals will not be identifiable.[10] An example is demand for a hypothetical product with characteristics outside their support in the data generating process. This kind of limitation is not special to our setting, of course: extrapolation outside the support of the data generating process typically requires some parametric structure. Our results, however, provide conditions under which such structure will be necessary *only* for such extrapolation. Furthermore, one may have more confidence in out-of-sample extrapolations if the in-sample preferences are nonparametrically identified.

A second issue more special to the demand application concerns welfare. Our specification of preferences has sufficient structure to characterize changes in the distribution of utilities in meaningful units and, therefore, of utilitarian social welfare. In particular, (5) incorporates

[10] The economic model enables identification of some out-of-sample counterfactuals—for example, removal of a product from the choice set.

quasilinearity preferences.[11] However, (5) lacks the structure required for welfare analysis that depends on the *distribution* of welfare changes. Identification of Pareto improvements, for example, will require additional restrictions enabling one to link an individual consumer's position in the distribution of utilities before a policy change to that after. This is because our model specifies a distribution of conditional indirect utilities, not a distribution of parameters whose realizations can be associated with a given individual. This points out a limitation of nonparametric random utility models as a theoretical foundation for some kinds of welfare analysis.

It should not be surprising that some welfare calculations require stronger assumptions than those necessary to identify demand, and our results will help to clarify which questions require these additional restrictions and which do not. Nonetheless, this will be an important limitation in some applications. An example of a model with sufficient structure to carry out the additional welfare analyses is the linear random coefficients model with independent choice-specific taste shocks. Because we show identifiability of a more general model, it should be possible (under additional assumptions) to project the identified joint distributions onto the space of linear random coefficients models.

3 Binary Choice with Exogenous Characteristics

Often one will want to allow for endogeneity of at least one component of x_{jt}. In applications to demand estimation, in particular, price will typically be an observed characteristic that is correlated with the unobserved "quality" ξ_{jt} through the optimizing behavior of sellers.[12] However, we begin with the simple case of binary choice with exogenous x_{jt}. This will illustrate

[11] The quasilinearity generally will not be in income, but one can describe changes in aggregate compensating/equivalent variation in units of the normalized marginal utility for $z_{ijt}^{(1)}$. Income (and/or price) will typically enter preferences through the function μ in (5). The potential nonlinearity of μ, combined with our inability to track indivuals' positions in the distributions of normalized utilities as the choice environment varies, prevents characterization of aggregate compensating variation or equivalent variation in income units.

[12] In the case of demand estimation with endogenous prices, identification arguments using control variates do not appear to be applicable in general. This is because in most models price is chosen by a firm that has observed all the cost and demand "shocks" in the model, not just its own demand shock ξ_{jt}. This violates the usual requirement that the endogenous right-hand-side variable be one-to-one with a scalar unobservable, conditional on observables (see, e.g., Imbens and Newey (2006)). An exception is the case of binary choice with no cost shocks. For binary response models, Blundell and Powell (2004) consider identification and estimation of a linear semiparametric model using a control function approach.

key elements of our approach and may be of independent interest.

Here we can drop the subscript j, with consumer i selecting choice 1 (i.e., $y_{i1} = 1$) whenever

$$z_{it}^{(1)} + \mu\left(x_t, \xi_t, z_{it}^{(2)}, \omega_{it}\right) > 0.$$

We consider identification under the following assumptions.

Assumption 2 $\xi_t \perp\!\!\!\perp (x_t, z_{it})$.

Assumption 3 *supp* $z_{it}^{(1)} | x_t, z_{it}^{(2)} = \mathbb{R}\ \forall x_t, z_{it}^{(2)}$.

Assumption 2 states that we consider the case of exogenous observables. This assumption is relaxed in the following section. A "large support" condition like Assumption 3 is common in the econometrics literature on nonparametric and semiparametric identification of discrete choice models (e.g., Manski (1985), Matzkin (1992), Matzkin (1993), Lewbel (2000)).[13] We relax this assumption in section 5, where the analysis will also clarify the role that the large support assumption plays in the results that maintain it. Here we show that Assumptions 1-3 are sufficient for identification.

Begin by fixing a value of $\mathbf{z}_{it}^{(2)}$, which can then be suppressed. Rewrite (5) as

$$u_{it} = z_{it}^{(1)} + \mu_{it} \tag{8}$$

where we have let $\mu_{it} = \mu(x_t, \xi_t, \omega_{it})$. Holding t fixed, all variation in μ_{it} is due to ω_{it}. Thus, $\mu_{it} \perp\!\!\!\perp z_{it}^{(1)}$ by Assumption 1. Since the observed conditional probability that a consumer chooses the outside good is given by

$$p_0(x_t, \mathrm{w}_{it}) = \Pr\left(\mu_{it} \leq -z_{it}^{(1)}\right)$$

we see that Assumption 3 guarantees that the distribution of $\mu_{it}|t$ (i.e., of μ_{it} in market t) is

[13] As usual, the support of $z_{it}^{(1)}$ need not equal the entire real line but need only cover the support of $\mu\left(x_t, \xi_t, z_{ijt}^{(2)}, \omega_{it}\right)$. We will nonetheless use the real line (real hyperplane below) for simplicity of exposition.

identified from variation in $z_{it}^{(1)}$ within market t. Denote this distribution by $F_{\mu_{it}|t}(\cdot)$. This argument can be repeated for all markets t.

In writing $\mu_{it}|t$, we condition on the values of x_t and ξ_t, although only the former is actually observed. However, once we have determined the distribution of $\mu_{it}|t$ for all t, we can recover the value of each ξ_t.

To see this, let

$$\delta_t = med\left[\mu_{it}|t\right] = med\left[\mu_{it}|x_t, \xi_t\right].$$

Given $F_{\mu_{it}|t}(\cdot)$, each δ_t is known. Under Assumption 1, we can write

$$\delta_t = D\left(x_t, \xi_t\right) \tag{9}$$

for some function D that is strictly increasing in its second argument. To show identification of D, for $\tau \in (0,1)$ let $\delta^\tau(x_t)$ denote the τth quantile of $\delta_t|x_t$ across markets. By the strict monotonicity of D in ξ_t, this quantile is unique. By (9) and the normalization of ξ_t

$$\delta^\tau(x_t) = D\left(x_t, \tau\right).$$

Since $\delta^\tau(x_t)$ is identified for all x_t and τ, D is identified on supp $x_t \times (0,1)$. With D known, each ξ_t is known as well.

Above we obtained identification of the distribution of $\mu_{it}|t$. Now we also have shown identifiability of the latent ξ_t associated with each market t. Thus, for any (x_t, ξ_t) in their support, we now have identification of

$$\begin{aligned} F_\mu\left(m|x_t, \xi_t\right) &= \Pr\left(\mu\left(x_t, \xi_t, \omega_{it}\right) \leq m|x_t, \xi_t\right) \\ &= F_{\mu_{it}|t}(m) \end{aligned}$$

for all $m \in \mathbb{R}$. With (8) this proves the following result.

Theorem 1 *Consider the binary choice setting with preferences given by (5). Under Assump-*

tions 1–3, the distribution of u_{it} *conditional on any* (x_t, ξ_t, z_{it}) *in their support is identified.*

Our argument involved two simple steps, each standard on its own. First, we showed that variation in $z_{it}^{(1)}$ *within each market* can be used to trace out the distribution of preferences across consumers within a market. It is in this step that the role of idiosyncratic variation in tastes is identified. Antecedents for this step include Matzkin (1992), Matzkin (1993), Lewbel (2000), and indeed this idea is used in analyzing identification of a wide range of qualitative response and selection models (e.g., Heckman and Honoré (1990), Athey and Haile (2002)). Second, we use variation in choice characteristics across markets (within *and* across markets in the case of multinomial choice) to decompose the nonstochastic variation in utilities across products into the variation due to observables and that due to the choice-specific unobservables ξ_{jt}. This idea has been used extensively in estimation of parametric multinomial choice demand models following Berry (1994), Berry, Levinsohn, and Pakes (1995), and Berry, Levinsohn, and Pakes (2004). This second step is essential once we allow the possibility of endogenous choice characteristics (e.g., correlation between price and ξ_{jt}), as will nearly always be necessary when one considers demand estimation. Our approach for the more general cases follows the same outline.

4 Multinomial Choice with Endogenous Characteristics

Here we consider multinomial choice, allowing endogeneity of choice characteristics. Let $\mathbf{x}_t = (x_{1t}, \ldots, x_{J_t t})$. We consider the following generalization of the large support assumption:

Assumption 4 *For all* $\mathcal{J}_t$*, supp* $\left\{z_{ijt}^{(1)}\right\}_{j=1,\ldots,J_t} \mid \left\{x_{jt}, z_{ijt}^{(2)}\right\}_{j=1,\ldots,J_t} = \mathbb{R}^{J_t}$.

This is a strong assumption requiring sufficient variation in $\left(z_{i1t}^{(1)}, \ldots, z_{iJ_t t}^{(1)}\right)$ to move choice probabilities through the entire unit simplex. Equivalent conditions are assumed in the prior work on multinomial choice by Matzkin (1993), Lewbel (2000), and Briesch, Chintagunta, and Matzkin (2005). Such an assumption provides a natural benchmark for exploring identifiability under ideal conditions. As discussed previously, however, we will also explore results that do not require this assumption in section 5.

Without the assumption $(x_{1t},\ldots,x_{J_tt}) \perp\!\!\!\perp \left(\xi_{1t},\ldots,\xi_{J_tt}\right)$, we will require instruments. To state our instrumental variables assumptions, let $x_{jt} = \left(x_{jt}^{(1)}, x_{jt}^{(2)}\right)$, where $x_{jt}^{(1)}$ denotes the endogenous characteristics. We then let $\mathrm{w}_{jt} \equiv \left(x_{jt}^{(2)}, \tilde{\mathrm{w}}_{jt}\right)$ denote the vector of instrumental variables. We will assume $x_{jt}^{(1)}$ is continuously distributed across j and t, with conditional density function $f_x\left(x_{jt}^{(1)}|\mathrm{w}_{jt}\right)$. We begin with the exclusion restriction.

Assumption 5 $\xi_{jt} \perp\!\!\!\perp (w_{jt}, z_{ijt})\, \forall j, t.$

The remaining IV condition we take from Chernozhukov and Hansen (2005). To state it we will need some notation. For simplicity, fix $x_{jt}^{(2)}$ in what follows, dropping it from the notation, so that x_{jt} now represents only the endogenous $x_{jt}^{(1)}$. Let

$$\delta_{jt} = D\left(x_{jt},\xi_{jt}\right) \equiv med\left[\mu\left(x_{jt},\xi_{jt},\omega_{it}\right)|x_{jt},\xi_{jt}\right]$$

and let $f_\delta\left(\cdot|x_{jt},\mathrm{w}_{jt}\right)$ denote its density conditional on w_{jt}.[14] Fix some small positive constants $\epsilon_q, \epsilon_f > 0$. For each $\tau \in (0,1)$, define $\mathcal{L}(\tau)$ to be the convex hull of functions $m(\cdot,\tau)$ that satisfy
(i) for all w_{jt}, $\Pr\left(\delta_{jt} \le m\left(x_{jt},\tau\right)|\tau,\mathrm{w}_{jt}\right) \in [\tau-\epsilon_q, \tau+\epsilon_q]$; and
(ii) for all x in the support of x_{jt}, $m(x,\tau) \in s_x \equiv \{\delta : f_\delta(\delta|x,\mathrm{w}) \ge \epsilon_f\ \forall \mathrm{w} \text{ with } f_x(x|\mathrm{w}) > 0\}$.

Assumption 6 *The random variables x_{jt} and δ_{jt} have bounded support. For any $\tau \in (0,1)$, for any bounded function $B(x,\tau) = m(x,\tau) - D(x,\tau)$ with $m(\cdot,\tau) \in \mathcal{L}(\tau)$ and $\varepsilon_{jt} \equiv \delta_{jt} - D(x_{jt},\tau)$, $E\left[B\left(x_{jt}\right)\psi\left(x_{jt},w_{jt}\right)|w_{jt}\right] = 0$ a.s. only if $B\left(x_{jt}\right) = 0$ a.s., where $\psi(x,w) = \int_0^1 f_\varepsilon\left(\sigma B(x)\,|x,w\right)\,d\sigma$.*

Assumption 6 is a particular type of "bounded completeness" condition, ensuring that the instruments induce sufficient variation in the endogenous variables. We take this conditions directly from Chernozhukov and Hansen (2005) (Appendix C).[15] This plays the role of the standard rank condition for linear models, but for the nonparametric nonseparable model $\delta = D(x,\xi)$. With these assumptions, we can prove the following result

[14] Chernozhukov and Hansen's "rank invariance" property holds here case because the same unobservable ξ_{jt} determines potential values of δ_{jt} for all possible values of the endogenous characteristics.
[15] They discuss sufficient conditions. We also consider an alternative to Assumption 6 below.

Theorem 2 *Under the representation of preferences in (5), suppose Assumptions 1, 4, 5, and 6 hold. Then the joint distribution of* $\{u_{ijt}\}_{j\in\mathcal{J}_t}$ *conditional on any* $\left(\mathcal{J}_t, \left\{(x_{jt}, z_{ijt}, \xi_{jt})\right\}_{j\in\mathcal{J}_t}\right)$ *in their support is identified.*

Proof. Fix $\mathcal{J}_t$, with $J_t = J$. Fix a value of the vector $\left(z^{(2)}_{i1t}, \ldots, z^{(2)}_{iJt}\right)$ and drop these arguments in what follows. Let $\mu_{ijt} = \mu\left(x_{jt}, \xi_{jt}, \omega_{it}\right)$. Observe that

$$\lim_{\substack{z^{(1)}_{ikt}\to-\infty \\ \forall k\neq j}} p_{ijt} = \Pr\left(z^{(1)}_{ijt} + \mu_{ijt} \geq 0\right).$$

Holding t fixed, $\mu_{ijt} \perp\!\!\!\perp z^{(1)}_{ijt}$ by Assumption 1. Assumption 4 then guarantees identification of the marginal distribution of each $\mu_{ijt}|t$. This implies identification of the conditional median

$$\begin{aligned} \delta_{jt} &= med\left[\mu\left(x_{jt}, \xi_{jt}, \omega_{it}\right)|t\right] \\ &\equiv D\left(x_{jt}, \xi_{jt}\right) \end{aligned} \tag{10}$$

where the unknown function D is strictly increasing in ξ_{jt}. Under Assumption 6, Theorem 4 of Chernozhukov and Hansen (2005) implies that D (and therefore each ξ_{jt}) is identified. Finally, observe that

$$\begin{aligned} p_{i0t} &= \Pr\left(z^{(1)}_{i1t} + \mu_{i1t} < 0, \ldots, z^{(1)}_{iJt} + \mu_{iJt} < 0\right) \\ &= \Pr\left(\mu_{i1t} < -z^{(1)}_{i1t}, \ldots, \mu_{iJt} < -z^{(1)}_{iJt}\right) \end{aligned} \tag{11}$$

so that Assumption 4 implies identification of the joint distribution of $(\mu_{i1t}, \ldots, \mu_{iJt})\,|t$. Since each x_{jt} is observed and ξ_{jt} is identified, this implies identification of the joint distribution of $(\mu_{i1t}, \ldots, \mu_{iJt})$ conditional on any $(x_{1t}, \xi_{1t}, z_{i1t}), \ldots, (x_{Jt}, \xi_{Jt}, z_{iJt})$ in their support given $\mathcal{J}_t$. Since $u_{ijt} = z^{(1)}_{ijt} + \mu_{ijt}$, the result follows. ■

This result demonstrates the identifiability of a very general model of multinomial choice with endogeneity. A possible limitation is that Assumption 6 is both difficult to check and difficult to interpret. Whether there are useful sufficient conditions on primitives delivering this property is an open question of broad interest in the literature on nonparametric instrumental variables regression, but beyond the scope of this paper. However, if we are willing to impose somewhat more structure on the utility function, we can obtain a more intuitive sufficient condition. Doing so will also enable us to relax the excludability restriction to require only mean independence.

To show this, suppose each consumer i's conditional indirect utilities can be represented by

$$\tilde{u}_{ijt} = \beta_{it} z_{ijt}^{(1)} + \tilde{\mu}\left(x_{jt}, z_{ijt}^{(2)}, \omega_{it}\right) + \gamma_{it}\xi_{jt} \qquad j = 1, \ldots, \mathcal{J}_t \tag{12}$$

where $\beta_{it} > 0$ w.p. 1, and the expectations $E\left[\beta_{it}\right]$, $E\left[\gamma_{it}\right]$, and $E\left[\tilde{\mu}\left(x_{jt}, z_{ijt}^{(2)}, \omega_{it}\right) | x_{jt}, z_{ijt}^{(2)}\right]$ are finite. This imposes a restriction relative to (4) but is still quite general relative to the prior literature. It is similar to the specification in Lewbel (2000), for example, but with random coefficients on both $z_{ijt}^{(1)}$ and ξ_{jt}, and with a nonparametric specification of $\tilde{\mu}$. A representation of preferences equivalent to (12) is

$$u_{ijt} = z_{ijt}^{(1)} + \mu\left(x_{jt}, \xi_{jt}, z_{ijt}^{(2)}, \omega_{it}\right) \qquad \forall i, j = 1, \ldots, \mathcal{J}_t \tag{13}$$

where now

$$\mu\left(x_{jt}, \xi_{jt}, z_{ijt}^{(2)}, \omega_{it}\right) = \frac{\tilde{\mu}\left(x_{jt}, z_{ijt}^{(2)}, \omega_{it}\right)}{\beta_{it}} + \frac{\gamma_{it}}{\beta_{it}}\xi_{jt}. \tag{14}$$

Let μ_{ijt} denote $\mu\left(x_{jt}, \xi_{jt}, z_{ijt}^{(2)}, \omega_{it}\right)$.

Here we will use a different normalization of ξ_{jt}. Instead of letting ξ_{jt} have a standard uniform distribution, we make the location normalization

$$E\left[\xi_{jt}\right] = 0$$

and scale normalization

$$E\left[\frac{\gamma_{it}}{\beta_{it}}\right] = 1. \tag{15}$$

Both are without further loss of generality. The latter defines units of the unobservable ξ_{jt} by fixing the mean marginal rate of substitution between $z_{ijt}^{(1)}$ and ξ_{jt}.

With this structure we can replace the full independence assumption with a mean independence assumption.

Assumption 7 $E\left[\xi_{jt}|\left(w_{jt}, z_{ijt}\right)\right] = 0\ \forall j, t, (w_{jt}, z_{ijt})$.

To prove identification of the joint distribution of $\{u_{ijt}\}_j$ conditional on $\left\{x_{jt}, z_{ijt}, \xi_{jt}\right\}_j$, first note that the argument in the proof of Theorem 2 remains valid here through equation (10). Recall that we have fixed the value of $\left(z_{i1t}^{(2)}, \ldots, z_{iJt}^{(2)}\right)$ and dropped these arguments. With the separable structure (14) and the normalization (15) now we let

$$\delta_{jt} \equiv E\left[\mu\left(x_{jt}, \xi_{jt}, \omega_{it}\right)|t\right] = D\left(x_{jt}\right) + \xi_{jt} \tag{16}$$

for some function D. As before, each δ_{jt} is identified from variation within each market. It is then straightforward to confirm that, under Assumption 7, the following "completeness" condition is equivalent to identification of the function D (Newey and Powell (2003)) from observation of $(\delta_{jt}, x_{jt}, \tilde{\mathrm{w}}_{jt})$.

Assumption 8 *For all functions $B\left(x_{jt}\right)$ with finite expectation, $E\left[B\left(x_{jt}\right)|w_{jt}\right] = 0$ a.s. implies $B\left(x_{jt}\right) = 0$ a.s.*

We can now state a second result for the multinomial choice model.

Theorem 3 *Under the utility representation (13), suppose Assumptions 1, 4, and 7 hold. Then the joint distribution of $\{u_{ijt}\}_{j\in\mathcal{J}_t}$ conditional on any $\left(\mathcal{J}_t, \left\{\left(x_{jt}, z_{ijt}, \xi_{jt}\right)\right\}_{j\in\mathcal{J}_t}\right)$ in their support is identified if and only if Assumption 8 holds.*

Proof. From the preceding argument, under the completeness Assumption 8, we have identification of D and therefore of each ξ_{jt}. The remainder of the proof then follows that of Theorem 2 exactly, beginning with (11). ■

The completeness condition (Assumption 8) is the analog of the rank condition in linear models. It requires that variation in w_{ijt} induce sufficient variation in $x_{jt}^{(1)}$ to reveal $D\left(x_{jt}\right)$ at all points x_{jt}. Lehman and Romano (2005) give standard sufficient conditions. Severini and

Tripathi (2006) point out that this condition is equivalent to the following: for any bounded function $f(x_{jt})$ such that $E[\ f(x_{jt})] = 0$ and $var(f(x_{jt})) > 0$, there exists a function $h(\cdot)$ such that $f(x_{jt})$ and $h(\mathrm{w}_{jt})$ are correlated. Additional intuition can be gained from the discrete case: as shown by Newey and Powell (2003), when x_{jt} and w_{jt} have discrete support $\left(\hat{x}^1, \ldots, \hat{x}^K\right) \times \left(\hat{w}^1, \ldots, \hat{w}^L\right)$, completeness corresponds to a full rank condition on the matrix $\{\sigma_{kl}\}$ where $\sigma_{kl} = \Pr(x_{jt} = \hat{x}^k | \mathrm{w}_{jt} = \hat{w}^l)$.

5 Limited Support

The large support assumption (Assumption 4) in the preceding section is both common in the literature and controversial. Our results using this condition demonstrate that sufficient variation in the vector $\left(z_{i1}^{(1)}, \ldots, z_{iJ_t}^{(1)}\right)$ can identify the joint distribution of utilities on their full support. Although our results describe only sufficient conditions for identifiability, it should not be surprising that a large support assumption may be needed: if the observable data can move choice probabilities only through a subset of the unit simplex, we should only hope to identify the joint distribution of utilities on a subset of *their* support. Of course, an important question is whether even these more limited hopes are fulfilled. In particular, one would like to understand how heavily the results rely on the tails of the large support, and to know what can be learned from more limited variation in the data. We explore these questions here.

We show that much more limited variation can be sufficient to identify the structural choice probabilities $\rho_j\left(\mathcal{J}_t, \{x_{jt}, \xi_{jt}, z_{ijt}\}_{j \in \mathcal{J}_t}\right)$ at all points of support. The relaxed condition of a "common choice probability" requires that there be vector of choice probabilities that is attainable in every market when conditioning on the appropriate vector $\left(z_{i1t}^{(1)}, \ldots, z_{iJ_tt}^{(1)}\right)$ for each market t. As discussed above, the structural choice probabilities are sufficient by themselves for many questions that motivate estimation of discrete choice demand models. Given the observability of p_{ijt}, the essential step is demonstrating identifiability of the choice-specific unobservables ξ_{jt}.

In addition, we show a type of continuity that suggests that the hopes described above are fulfilled. In particular, there is a natural sense in which moving from our limited support condition to the full support condition moves the identified features of the model smoothly toward the full identification results of the preceding section. For the multinomial choice case we obtain this result under a somewhat more restrictive specification of preferences. Up to this qualification, however, this should be a comforting observation regarding the robustness of the results we obtain using the large support assumption. Although we can obtain full identifiability of the model only with the large support, the result is not knife-edge. In particular, the tails of the support are needed only to determine the tails of the joint distributions of utilities.

In section 5.1 we first explore identifiability of the structural choice probabilities with no support assumption whatsoever (indeed even without separability in, or even existence of, $z_{ijt}^{(1)}$). Thus far we have obtained positive results only for the binary choice case in this situation. This offers one motivation for exploration in section 5.2 of conditions that rely on the separable structure used above but utilizing the common choice probability condition. Our results under this condition will illustrate the robustness discussed above.

5.1 Identification of ξ_{jt} With No Support Condition

Consider the general specification of preferences in (1), without requiring separability in $z_{it}^{(1)}$. In the case of binary choice the consumer selects the inside good if

$$u\left(x_t, \xi_t, z_{it}, \omega_{it}\right) > 0. \tag{17}$$

Note that we have dropped the earlier requirement of additive separability in $z_{it}^{(1)}$. In fact, there need not exist any individual-choice specific observables at all. Under Assumption 1, the probability of the event (17) can be written

$$\pi_{it} = \rho\left(x_t, z_{it}, \xi_t\right) \tag{18}$$

where ρ is a strictly increasing function of ξ_t. Since these probabilities are observed (along with x_t, z_{it}) the results of Chernozhukov and Hansen (2005) can be applied as above to identify the function ρ and, therefore, each latent ξ_t. With each ξ_t known, the structural choice probabilities $\rho\left(x_t, z_{it}, \xi_t\right)$ are then identified at all points of support.

Theorem 4 *Consider the binary choice model with the representation of preferences in (1). Suppose Assumptions 1, 5, and 6 hold. Then the structural choice probabilities $\rho(x_t, z_{it}, \xi_t)$ are identified at all points of support.*

A key to this result can be seen from the case without endogeneity. For that case, consider the cumulative distribution $F_\pi(\cdot|x_t, z_{it})$ of π_{it} across markets conditional on (x_t, z_{it}). The market with choice probability at the τth quantile of the distribution of $F_\pi(\cdot|x_t, z_{it})$ is the market with $\xi_t = \tau$ (recalling that ξ_t is $u(0,1)$). Thus, one identifies each ξ_t by inverting equation (18):

$$\xi_t = \rho^{-1}(\pi_{it}; x_t, z_{it}) = F_\pi(\pi_{it}|x_t, z_{it}).$$

The instrumental variables results of Chernozhukov and Hansen (2005) enable us to extend this idea to the case of endogenous characteristics.

An open question is whether a similar possibility exists in the multinomial case. Letting π_{ijt} denote the probability that an individual i in market t selects good j, we have

$$\pi_{ijt} = \rho_j(\mathcal{J}_t, x_{1t}, \ldots, x_{Jt}, z_{i1t}, \ldots z_{iJt}, \xi_{1t}, \ldots, \xi_{Jt}) \quad j = 1 \ldots J. \tag{19}$$

With $(\pi_{ijt}, \mathcal{J}_t, x_{1t}, \ldots, x_{Jt}, z_{i1t}, \ldots z_{iJt})$ observable, this yields, for each market t, a system of J_t equations in the J_t unknown values of ξ_{jt}. Following the "inversion" result of Berry (1994) and Berry and Pakes (2007), we can solve for the product-level unobservables in terms of the purchase probabilities:

$$\xi_{jt} = \rho_j^{-1}(\pi_{1t}, \ldots, \pi_{Jt}; \mathcal{J}_t, x_{1t}, \ldots, x_{Jt}, z_{i1t}, \ldots z_{iJt}).$$

5.2 Identification of ξ_{jt} with a Common Choice Probability

5.2.1 Binary Choice

As before, we will begin with the binary choice case to illustrate our insights most simply. Here we will discuss two results. We first consider the general specification of preferences in (5) above. We then consider the a more restrictive specification (analogous to (13)) that appears to be more useful for the multinomial case. For both results we make the following "common choice probability" assumption:

Assumption 9 *For some* $\tau \in (0,1)$*, for every market* t *there exists a unique* $z_t^\tau \in$*supp* $z_{it}^{(1)}$ *such that* $\Pr\left(y_{it} = 1 | z_{it}^{(1)} = z_t^\tau\right) = \tau$.

Here we require sufficient variation in $z_{it}^{(1)}$ to push the choice probability to τ in each market, not over the whole interval $(0,1)$ in each market.[16] This is a much weaker requirement and is likely to be satisfied in many applications.

General Case Consider the specification of preferences in (5). In the binary choice case, the consumer chooses the inside good if (fixing $z_{it}^{(2)}$)

$$z_{it}^{(1)} + \mu\left(x_t, \xi_t, \omega_{it}\right) > 0.$$

With the common choice probability assumption, for each market t we can identify the value z_{it}^τ such that

$$\Pr\left(-\mu\left(x_t, \xi_t, \omega_{it}\right) < z_{it}^{(1)} | x_t, \xi_t, z_{it}^{(1)}\right)\Big|_{z_{it}^{(1)} = z_t^\tau} = \tau.$$

Then each z_t^τ is the τth quantile of the random variable $-\mu_{it} \equiv -\mu\left(x_t, \xi_t, \omega_{it}\right)$ conditional on t, i.e., on (x_t, ξ_t). Thus, we can write

$$z_t^\tau = \zeta\left(x_t, \xi_t; \tau\right) \tag{20}$$

for some function $\zeta\left(\cdot;\tau\right)$ that is strictly decreasing in ξ_t.

Identification of the function $\zeta\left(\cdot;\tau\right)$ and, therefore, of each ξ_t, then follows from (20) as in the preceding sections, again using the nonparametric instrumental variables result of Chernozhukov and Hansen (2005). With each ξ_t known, the observable choice probabilities reveal

[16] Implicitly we also require a continuous (region of) support for $\mu\left(x_t, \xi_t, z_{it}^{(2)}, \omega_{it}\right) | x_t, \xi_t, z_{it}^{(2)}$ to gaurantee uniqueness.

the structural choice probabilities

$$\rho\left(x_t, \xi_t, z_{it}\right) = \Pr\left(y_{it} = 1 | x_t, \xi_t, z_{it}\right) \tag{21}$$

at all points (x_t, ξ_t, z_{it}) of support. Thus, we have shown the following result.

Theorem 5 *In the binary choice model with preferences given by (5), suppose Assumptions 1, 5, 6, and 9 hold. Then the structural choice probabilities* $\rho\left(x_t, \xi_t, z_{it}\right)$ *are identified at all points* (x_t, ξ_t, z_{it}) *in their support.*

Although we cannot identify the full probability distribution of $u_{i1t}|x_t, z_{it}, \xi_t$, we can obtain some information about this distribution from every common choice probability. In particular, we can identify a $\zeta\left(\cdot;\tau\right)$ for each common choice probability τ, each then determining the τth quantile of $-\mu\left(x_t, \xi_t, \omega_{it}\right)$. Since

$$u_{it} = z_{it}^{(1)} + \mu\left(x_t, \xi_t, \omega_{it}\right)$$

this determines the corresponding quantiles of the distribution of u_{it} conditional on (x_t, ξ_t, z_{it}). In the limit—i.e., with sufficient variation in $z_{it}^{(1)}$ to make every $\tau \in (0,1)$ a common choice probability—all quantiles of the distribution of u_{it} conditional on (x_t, ξ_t, z_{it}) are identified, and we are back to full identification as in Theorem 2. This illustrates the notion of "continuity" discussed above: the tails of $z_{ijt}^{(1)}$ under the large support assumption are used only to identify the tails of the conditional distributions of utilities.

Additive ξ As before, we can replace Assumptions 5 and 6 with Assumptions 7 and 8 if we impose linear separability in ξ_t, as in (16). Here we also impose a further restriction on the utility specification in (12)—in particular, that $\frac{\gamma_{it}}{\beta_{it}}$ to equal one *for each consumer* rather than merely *in expectation.* After normalizing by γ_{it}, this leads to the representation

$$u_{ijt} = z_{ijt}^{(1)} + \mu\left(x_{jt}, z_{ijt}^{(2)}, \omega_{it}\right) + \xi_{jt} \qquad \forall i, j = 1, \ldots, \mathcal{J}_t. \tag{22}$$

This involves a significant restriction on preferences relative to our previous results, although a similar restriction has been required for the most general prior results for identification of linear semiparametric models with heteroskedasticity and endogeneity (e.g., Lewbel (2000)). Thus, our most restrictive specification (22) is still a generalization relative to the literature, even without our relaxation of the the large support assumption here.

Theorem 6 *In the binary choice model with preferences given by (22), suppose Assumptions 1, 7, 8, and 9 hold. Then the structural choice probabilities $\rho(x_t, \xi_t, z_{it})$ are identified at all points (x_t, ξ_t, z_{it}) in their support.*

Proof. Fixing $\mathbf{z}_{it}^{(2)}$, suppressing it, and defining z_t^τ as before, we have

$$\tau = \Pr(-\mu(x_t, \omega_{it}) < z_t^\tau + \xi_t \; |t)$$

Thus $z_t^\tau + \xi_t$ is the τth quantile of $-\mu(x_t, \omega_{it})$ conditional on t. Since quantiles of $\mu(x_t, \omega_{it})|t$ are functions of x_t alone, we can write

$$z_t^\tau + \xi_t = \zeta(x_t; \tau)$$

for some function $\zeta(\cdot; \tau)$, which gives

$$z_t^\tau = \zeta(x_t; \tau) - \xi_t. \tag{23}$$

Equation (23) can then be used as before to identify $\zeta(\cdot; \tau)$ (and therefore ξ_t) using the results of Newey and Powell (2003). As in the preceding section, this is sufficient to determine the structural choice probabilities $\rho(x, \xi_t, z_{it}) = \Pr(y_{it} = 1 | x_t, \xi_t, z_{it})$ at all points (x_t, ξ_t, z_{it}) of support. ■

Once we have identified ξ_t from some common choice probability τ, we can generate a large set of quantiles of $\mu(x_t, \omega_{it})$, even if τ is the only common choice probability. Consider all the choice probabilities q generated within market t by moving $z_{it}^{(1)}$ across its *entire support* in t. Call each of these $q\left(z_{it}^{(1)}\right)$. For each $\left(z_{it}^{(1)}, q\left(z_{it}^{(1)}\right)\right)$ pair, the $q\left(z_{it}^{(1)}\right)$th quantile of $\mu(x_t, \omega_{it})$ is given by

$$\zeta\left(x_t; q\left(z_{it}^{(1)}\right)\right) = z_{it}^{(1)} + \xi_t.$$

Thus, we obtain a new quantile of $\mu(x_t, \omega_{it})$ for every observed choice probability in market t. Additional quantiles of $\mu(x_t, \omega_{it})$ may also be obtained from other markets t' with $x_{t'} = x_t$ but $\xi_{t'} \neq \xi_t$. Thus we may be able to recover the CDF of $u_{it}|x_t, \xi_t, z_{it}$ over a significant portion of its domain, even with only a single common choice probability.

5.2.2 Multinomial Choice

For the multinomial case we will maintain the more restrictive representation of preferences in (22), where

$$u_{ijt} = z_{ijt}^{(1)} + \mu\left(x_{jt}, z_{ijt}^{(2)}, \omega_{it}\right) + \xi_{jt} \qquad \forall i, j = 1, \ldots, \mathcal{J}_t. \tag{24}$$

Letting $\triangle^{J_t}$ denote the $|\mathcal{J}_t|-1$ dimensional unit simplex, we generalize the previous common choice probability assumption in the natural way:

Assumption 10 *For all $\mathcal{J}_t$, there exists some $q = (q_0, q_1, \ldots, q_J) \in \triangle^{J_t}$ such that for every market t there is a unique vector $\mathbf{z}_t^q = (z_{1t}^q, \ldots, z_{1t}^q) \in supp\left(z_{i1t}^{(1)}, \ldots, z_{iJ_tt}^{(1)}\right)$ such that $q_j = \Pr(y_{it} = j \mid x_{1t}, \ldots, x_{J_tt}, z_{i1t}, \ldots, z_{iJ_tt})_{\mathbf{z}_{it}^{(1)} = \mathbf{z}_t^q}$ for all $j = 1, \ldots, J_t$.*

If $\mu\left(x_{jt}, z_{ijt}^{(2)}, \omega_{it}\right)$ is continuously distributed, uniqueness of $\mathbf{z}_t^q$ is guaranteed by the one-to-one mapping between choice probabilities and the deterministic component of utilities, demonstrated in Berry (1994) and Berry and Pakes (2007). Beyond this, the requirement of Assumption 10 is that the vector $\left(z_{i1t}^{(1)}, \ldots, z_{iJt}^{(1)}\right)$ have sufficient support to drive the choice probability vector to q in each market. This is clearly weaker than the full support condition, which requires all elements of $\triangle^{J_t}$ to be common choice probabilities. In practice, this condition is most likely to hold when the choice characteristics $\left(x_{jt}, \xi_{jt}\right)$ do not vary too wildly across markets.

Theorem 7 *In the multinomial choice model with preferences given by (24), suppose Assumptions 1, 7 , 8, and 10 hold. Then the structural choice probabilities $\rho_j\left(\mathcal{J}_t, \{x_{jt}, \xi_{jt}, z_{ijt}\}_{j\in\mathcal{J}_t}\right)$ are identified at all $\left(\mathcal{J}_t, \{x_{jt}, \xi_{jt}, z_{ijt}\}_{j\in\mathcal{J}_t}\right)$ in their support.*

Proof. Under (24), choice probabilities depend only on the sums

$$\lambda_{ijt} \equiv z_{ijt}^{(1)} + \xi_{jt} \tag{25}$$

rather than on each $z_{ijt}^{(1)}$ and ξ_{jt} separately. In particular, fixing the vector $\mathbf{z}_t^{(2)}$,

$$\begin{aligned} p_{ijt} &= \Pr(y_{it} = j \mid x_{1t}, \ldots, x_{J_t t}, \xi_{1t}, \ldots, \xi_{J_t t}, z_{i1t}, \ldots, z_{iJ_t t}) \\ &= \Pr(y_{it} = j \mid x_{1t}, \ldots, x_{J_t t}, \lambda_{i1t}, \ldots, \lambda_{iJ_t t}) \\ &= \Pr\left(\mu\left(x_{jt}, \omega_{it}\right) + \lambda_{ijt} \geq \max\left\{0, \max_k \ \mu\left(x_{kt}, \omega_{it}\right) + \lambda_{ikt}\right\}\right). \end{aligned}$$

From (25) and Assumption 10, for all $(x_{1t}, \ldots, x_{J_t t})$ there is a unique vector

$$\lambda\left(\mathbf{x}_t, q\right) = \left(\lambda_1\left(\mathbf{x}_t, q\right) ., \ldots, \lambda_{J_t}\left(\mathbf{x}_t, q\right)\right)$$

such that

$$\lambda_j\left(\mathbf{x}_t, q\right) = \xi_{jt} + z_{jt}^q \tag{26}$$

and

$$q_j = p_{ijt} = \Pr(y_{it} = j \mid x_{1t}, \ldots, x_{J_t t}, \ \lambda_1\left(\mathbf{x}_t, q\right), \ldots, \lambda_{J_t}\left(\mathbf{x}_t, q\right)) \quad \forall j.$$

From (26),

$$z_{jt}^q = \lambda_j\left(\mathbf{x}_t, q\right) - \xi_{jt} \quad \forall j, t. \tag{27}$$

These equations identify the functions $\lambda_j\left(\cdot, q\right)$ and each ξ_{jt} for all j and t under Assumptions 7 and 8, using the results for the additively separable nonparametric IV regression in Newey and Powell (2003). As demonstrated above, knowledge of all ξ_{jt} identifies the structural choice probability functions. ■

6 Testable Restrictions

Our model is quite general but relies on two important assumptions: (i) existence of a vertical additively separable observable, $z_{ijt}^{(1)}$; (ii) adequacy of a scalar vertical choice-specific unobservable, ξ_{jt}. Here we show that both assumptions imply testable restrictions.[17]

Our assumption that preferences can be represented by conditional indirect utilities in which $z_{ijt}^{(1)}$ enters in an additively separable fashion (with positive coefficient) has two implications. The first concerns the reduced form choice probabilities, and is immediate from the assumption that the utility from good j is strictly increasing in $z_{ijt}^{(1)}$.

Theorem 8 *Suppose preferences are characterized by* (5)*. Then under Assumption 1,* $p_{ijt} \equiv \Pr\left(y_{it} = j | \mathcal{J}_t, \{x_{jt}, w_{jt}, z_{ikt}\}_{k \in \mathcal{J}_t}\right)$ *is increasing in* $z_{ijt}^{(1)}$*.*

The second involves an overidentifying restriction. Let $F_u\left(\cdot | \{x_{jt}, \xi_{jt}, z_{ijt}\}_{j \in \mathcal{J}_t}\right)$ denote the joint distribution of conditional indirect utilities for the choice set $\mathcal{J}_t$. Define the sets

$$A_j = \left\{ (u_1, \ldots, u_{J_t}) \in \mathbb{R}^{J_t} : u_j > \max\left\{0, \max_{k \neq j} \ u_k \right\} \right\}.$$

so that

$$\rho_j\left(\mathcal{J}_t, \{x_{jt}, \xi_{jt}, z_{ijt}\}_{j \in \mathcal{J}_t}\right) = \int_{A_j} dF_u\left(u_1, \ldots, u_{J_t} | \{x_{jt}, \xi_{jt}, z_{ijt}\}_{j \in \mathcal{J}_t}\right) \quad \forall j. \tag{28}$$

Under the large support condition, we showed in Theorems 2 and 3 that $F_u\left(\cdot | \{x_{jt}, \xi_{jt}, z_{ijt}\}_{j \in \mathcal{J}_t}\right)$ was identified, as was each ξ_{jt}. Knowledge of $F_u\left(\cdot | \{x_{jt}, \xi_{jt}, z_{ijt}\}_{j \in \mathcal{J}_t}\right)$ determines the right-hand-side of (28). With each ξ_{jt} known, the observable choice probabilities also directly identify the structural choice probabilities $\rho_j\left(\mathcal{J}_t, \{x_{jt}, \xi_{jt}, z_{ijt}\}_{j \in \mathcal{J}_t}\right)$ on the left-hand side, as noted above. Noting that the proofs of Theorems 2 and 3 did not use the condition (28) (but did rely on the linearity of utilities in $z_{ijt}^{(1)}$), we have the following.

Theorem 9 *Under the hypotheses of Theorem 2 or Theorem 3, the overidentifying restrictions* (28) *must hold.*

[17] The random utility discrete choice paradigm with stable preferences (as in our Assumption 1) also generates well known testable restrictions (see, e.g., Block and Marschak (1960) and Falmagne (1978)).

The assumption of a scalar vertical unobservable also leads to testable implications. We show this here for the binary choice case for simplicity. To state the result it will be useful to recall Theorem 5 and let $\xi_t\left(z_t^\tau;\tau,x_t\right)$ denote the value of ξ_t identified from the common choice probability τ in each market t. As usual, we condition on $z_{it}^{(2)}$ and suppress it in the notation.

Theorem 10 *In the binary choice model with preferences given by (5), suppose Assumptions 1, 5, 6, and 9 hold. Then $\xi_t\left(z_t^\tau;\tau,x_t\right)$ must be strictly decreasing in z_t^τ across markets.*

Proof. This is immediate from the fact that u_{it} is strictly increasing in both $z_{it}^{(1)}$ and ξ_t under the assumptions of the model. ■

The following example shows one way that a model with a horizontal rather than a vertical unobservable characteristic can lead to a violation of this restriction.

Example 5 *Suppose $\mu\left(x_t,\xi_t,\phi_{it}\right)=-\nu_{it}\xi_t$, with $\nu_{it}\sim N(0,1)$. Take $\tau>1/2$ and consider the set of markets in which $\xi_t\left(z_t^\tau;\tau,x_t\right)>0$.*[18] *Recall that each z_t^τ is observable and defined such that $\Pr\left(\nu_{it}\xi_t<z_t^\tau\right)=\tau$. Letting Φ denote the standard normal CDF, this requires*

$$\Phi\left(\frac{z_t^\tau}{\xi_t}\right)=\tau\quad\forall t. \tag{29}$$

Therefore, by construction, $\frac{z_t^\tau}{\xi_t}$ will take the same value in every market. Since each z_t^τ must also be positive when $\tau>1/2$, this requires a strictly positive correspondence between z_t^τ and ξ_t across markets. Thus, $\xi_t\left(z_t^\tau;\tau,x_t\right)$ will violate the testable restriction of the Theorem 10.

Theorem 11 *In the binary choice model with preferences given by (5), suppose Assumptions 1, 5, 6, and 9 hold. In addition, suppose that for distinct τ and τ' in the interval $(0,1)$, for every market t there exists a unique $z_t^\tau\in supp\ z_{it}^{(1)}$ such that $\Pr\left(y_{it}=1|z_{it}^{(1)}=z_t^\tau\right)=\tau$ and a unique $z_t^{\tau'}\in supp\ z_{it}^{(1)}$ such that $\Pr\left(y_{it}=1|z_{it}^{(1)}=z_t^{\tau'}\right)=\tau'$. Then $\xi_t\left(z_t^\tau;\tau,x_t\right)=\xi_t\left(z_t^{\tau'};\tau',x_t\right)$ for all t.*

Proof. This is immediate from the fact that, under the assumptions of the model, $\xi_t\left(z_t^\tau;\tau,x_t\right)=\xi_t\left(z_t^{\tau'};\tau',x_t\right)=\xi_t$. ■

[18] An analogous argument applies to the set of markets with $\xi_t\left(z_t^\tau;\tau,x_t\right)<0$.

The following example demonstrates that this restriction can fail if the restriction to a scalar unobservable is violated.

Example 6 *Consider a model with two vertical unobservables, ξ_t^1 and ξ_t^2. Let*

$$\mu\left(x_t, \xi_t^1, \xi_t^2, \omega_{it}\right) = \begin{cases} \nu_{it}\left(\xi_t^1 + \xi_t^2\right) & \nu_{it} < 1/2 \\ \nu_{it}\left(\xi_t^1 + 2\xi_t^2\right) & \nu_{it} \geq 1/2 \end{cases}$$

with $\nu_{it} \sim u[0,1]$. Let ξ_t^1 and ξ_t^2 be independent, each uniform on $(0,1)$. By definition, when $z_{it}^{(1)} = z_t^\tau$ only consumers with $v_{it} > 1-\tau$ choose the inside good. Thus, the value of z_t^τ is determined by the preferences of the consumer with $v_{it} = 1-\tau$. Now consider the $\xi_t(\tau)$ inferred under the incorrect assumption of a scalar unobservable. When $\tau > 1/2$, $\xi_t(\tau) = F_{\xi^1+\xi^2}\left(\xi_t^1 + \xi_t^2\right)$ where $F_{\xi^1+\xi^2}$ is the CDF of the sum of two independent uniform random variables. Thus, if for market t, $\left(\xi_t^1 + \xi_t^2\right)$ falls at the σ quantile in the cross-section of markets, $\xi_t(\tau)$ will equal σ. For $\tau' < 1/2$, $\xi_t(\tau') = F^{-1}_{\xi^1+2\xi^2}\left(\xi_t^1 + 2\xi_t^2\right)$; i.e, if $\xi_t^1 + 2\xi_t^2$ fall at the σ' quantile of this sum in the cross section of markets, $\xi_t(\tau')$ will be σ'. In general, $\sigma \neq \sigma'$.

7 Relation to the Literature

Important early work on identification of discrete choice models includes Manski (1985), Manski (1988), Matzkin (1992), and Matzkin (1993). Manski considered a semi-parametric linear random coefficients model of binary response, focusing on identification of the slope parameters determining mean utilities. Matzkin considered nonparametric specifications of binomial and multinomial response models with independent, additively separable taste shocks (no random coefficients). None of this earlier work allowed choice-specific unobservables ξ_{jt} or endogenous choice characteristics.

Relative to this early work our framework involves two important generalizations. One is heterogeneity in consumer preferences for choice characteristics. The other is the existence of unobserved choice characteristics that are correlated with observable choice characteristics.

Heterogeneity in preferences for characteristics has previously been explored using random coefficients models. Identification of linear random-coefficients binary choice models has been considered by Ichimura and Thompson (1998) and Gautier and Kitamura (2007). Briesch, Chintagunta, and Matzkin (2005) consider multinomial choice, allowing some generalizations of the linear random coefficients model but requiring mutual independence of all taste shocks. Recent work by Fox and Gandhi (2008) explores identifiability of several related models, including a semiparametric random coefficients multinomial choice model. All of these papers require additional assumptions, and none relaxes our requirement of linearity in at least one characteristic. Importantly, none provides a complete treatment of choice-specific unobservables or endogenous choice characteristics, which are typically essential in the context of demand estimation.

Work considering endogenous choice characteristics includes Lewbel (2000), Honoré and Lewbel (2002), Hong and Tamer (2004), Blundell and Powell (2004), Lewbel (2005), and Magnac and Maurin (2007). These all consider linear semiparametric models with a single additively separable taste shock for each choice; i.e. no heterogeneity in preferences for choice characteristics. All but the two papers by Lewbel limit attention to binary response models, which can be significantly simpler and can, under additional restrictions, be amenable to control function methods in the context of demand estimation.

8 Conclusion

We have examined the identifiability of multinomial choice models of demand. Our framework allows a nonparametric distribution-free specification of random utility with endogenous choice characteristics, unknown heteroskedasticity, and correlated taste shocks. Our results are obtained using standard assumptions from the literature, and we have shown that for some questions a "large support" assumption can be relaxed considerably while preserving identification of key features. Our general approach relied on using variation in choice probabilities within a market to trace out some or all of the distribution of utilities conditional on a choice set, while using variation in the marginal distribution of utility across choices (within and across markets) to identify the role of unobserved product characteristics.

One reason we have been able to make progress in well worn territory is the generality of the model we consider. This may be counterintuitive. One might have thought that starting with more restrictive models—e.g., the linear random coefficient model—would make identification results more easily attainable. So it might be surprising that we obtain results for a much more general model while simultaneously avoiding some restrictions (e.g., independence of random coefficients) required in the prior literature. However, this intuition may be misleading. To see why, consider the dimensionality of the primitives of our model. Given a choice set, the joint distribution of the conditional indirect utilities is J_t-dimensional. Contrast this with a linear random coefficients model without unobserved product characteristics, where

$$u_{ijt} = x_{jt}\beta_i + \epsilon_{ijt}.$$

In that case the structure consists of the joint distribution of $\{\beta_i, \epsilon_{ij}\}_j$. If x_{jt} is k-dimensional, the joint distribution of $\{\beta_i, \epsilon_{ij}\}_j$ has $J_t + k$ components. Since the observable (conditional on the choice set) is the J_t-dimensional vector $(p_{i1t}, \ldots, p_{iJ_tt})$, it should not be surprising that additional assumptions must be imposed to obtain identification of the linear random coefficients model. Focusing on the nonparametric joint distribution of utilities *directly,* however, naturally limits the dimensionality of the primitives to that of the observables.

A second source of progress is our recognition that for many purposes the structural choice probabilities alone determine the answers to the economic questions of interests. In particular, excluding welfare analysis, motivations for demand estimation in practice typically require identification only of the demand system, not of the underlying preference structure. For this, the critical step is uncovering choice-specific unobservables, ξ_{jt}. Such unobservables have often been ignored in the prior literature, limiting the applicability to many important environments (notably, almost any application to demand). We have shown that these choice-specific unobservables, and therefore the structural choice probabilities that fully characterize demand, are identified with under support conditions likely to hold in many applications.

One important distinction between our work and much (indeed, most) of the prior literature is our neglect of estimation. Although our identification proofs may suggest new nonparametric or semiparametric estimation approaches, additional work is needed. Moving from identification typically requires additional smoothness conditions at a minimum, and in practice additional structure may be desirable for estimation. However, our main objective has been to provide additional exploration of identification in a very general setting in order to better understand the potential scope and limitations of empirical work using choice data to estimate the underlying structure of demand and preferences.

References

Anderson, S., A. DePalma, and F. Thisse (1992): *Discrete Choice Theory of Product Differentiation*. MIT Press, Cambridge MA.

Athey, S., and P. A. Haile (2002): "Identification of Standard Auction Models," *Econometrica*, 70(6), 2107–2140.

Athey, S., and G. W. Imbens (2007): "Discrete Choice Models with Multiple Unobserved Choice Characteristics," *International Economic Review*, 48, 1159–1192.

Bayer, P., F. Ferreira, and R. McMillan (2007): "A Unified Framework for Measuring Preferences for Schools and Neighborhoods," *Journal of Political EconomyReview of Economic Studies*, 115(5), 588–638.

Berry, S. (1994): "Estimating Discrete Choice Models of Product Differentiation," *RAND Journal of Economics*, 23(2), 242–262.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 60(4), 889–917.

——— (2004): "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Vehicle Market," *Journal of Political Economy*, 112(1), 68–105.

Berry, S. T., and P. A. Haile (2008): "Identification of Discrete Choice Demand Models from Market Level Data," Discussion paper, Yale University.

BERRY, S. T., AND A. PAKES (2007): "The Pure Characteristics Demand Model," Discussion paper, Yale University.

BLOCK, H., AND J. MARSCHAK (1960): "Random Orderings and Stochastic Theories of Responses," in *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, ed. by I. Olkin, S. Ghurye, W. Hoeffding, W. G. Mado, and H. B. Mann. Stanford University Press.

BLUNDELL, R. W., AND J. L. POWELL (2004): "Endogeneity in Semiparametric Binary Response Models," *Review of Economic Studies*, 71, 655–679.

BRESNAHAN, T. (1981): "Departures from Marginal Cost Pricing in the American Automobile Industry," *Journal of Econometrics*, 17, 201–227.

BRIESCH, R. A., P. K. CHINTAGUNTA, AND R. L. MATZKIN (2005): "Nonparametric Discrete Choice Models with Unobserved Heterogeneity," Discussion paper, Northwestern University.

CAPPS, C., D. DRANOVE, AND M. SATTERTHWAITE (2003): "Competition and Market Power in Option Demand Markets," *RAND Journal of Economics*, 34(5), 737–763.

CHERNOZHUKOV, V., AND C. HANSEN (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73(1), 245–261.

DOMENICH, T., AND D. MCFADDEN (1975): *Urban Travel Demand: A Behavioral Analysis*. North Holland, Amsterdam.

FALMAGNE, J.-C. (1978): "A Representation Theorem for Finite Random Scale Systems," *Journal of Mathematical Psychology*, 18, 52–72.

FOX, J., AND A. GANDHI (2008): "Identifying Selection and Discrete-Continuous Models Using Mixtures," Discussion paper, University of Chicago.

GAUTIER, E., AND Y. KITAMURA (2007): "Nonparametric Estimation in Random Coefficients Binary Choice Models," Discussion paper, Yale.

Goldberg, P. K. (1995): "Product Differentiation and Oligopoly in International Markets: The Case of the U.S. Automobile Industry," *Econometrica*, 63(4), 891–951.

Guadagni, P. M., and J. D. C. Little (1983): "A Logit Model of Brand Choice Calibrated on Scanner Data," *Marketing Science*, 2(3), 203–238.

Hastings, J., D. Staiger, and T. Kane (2007): "Preferences and Heterogeneous Treatment Effects in a Public School Choice Lottery," Discussion paper, Yale University.

Hausman, J. A. (1996): "Valuation of New Goods under Perfect and Imperfect Competitioin," in *The Economics of New Goods*, ed. by T. F. Bresnahan, and R. J. Gordon, chap. 5, pp. 209–248. University of Chicago Press, Chicago.

Heckman, J. J., and B. E. Honoré (1990): "The Empirical Content of the Roy Model," *Econometrica*, 58, 1121–1149.

Ho, K. (2007): "Insurer-Provider Networks in the Medical Care Market," Discussion paper, Columbia University.

Hong, H., and E. Tamer (2004): "Endogenous Binary Choice Mode with Median Restrictions," *Economics Letters*, pp. 219–224.

Honoré, B. E., and A. Lewbel (2002): "Semiparametric Binary Choice Panel Data Models Without Strictly Exogenous Regressors," *emet*, 70(5), 2053–2063.

Ichimura, H., and T. S. Thompson (1998): "Maximum Likelihood Estimation of a Binary Choice Model with Random Coefficients of Unknown Distribution," *Journal of Econometrics*, 86(2), 269–95.

Imbens, G., and W. Newey (2006): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," Discussion paper, M.I.T.

Kennan, J., and J. Walker (2006): "The Effect of Expected Income on Individual Migration Decisions," Discussion paper, University of Wisconsin-Madison.

Lehman, E., and J. P. Romano (2005): *Testing Statistical Hypotheses*. Springer, New York, 3 edn.

Lewbel, A. (2000): "Semiparametric Qualitative Response Model Estimation with Unknown Heteroscedasticity or Instrumental Variables," *Journal of Econometrics*, 97, 145–177.

——— (2005): "Simple Endogenous Binary Choice and Selection Panel Model Estimators," Discussion paper, Boston College.

Luce, R. D. (1959): *Individual Choice Behavior*. Wiley.

Magnac, T., and E. Maurin (2007): "Identification and Information in Monotone Binary Models," *Journal of Econometrics*, 139, 76–104.

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

——— (1985): "Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator," *Journal of Econometrics*, 27, 313–333.

——— (1988): "Identification of Binary Response Models," *Journal of the American Statitical Association*, 83(403), 729–738.

Matzkin, R. (1992): "Nonparametric and Distribution-Free Estimatoin of the Binary Choice and Threshold Crossing Models," *Econometrica*, 60(2).

——— (1993): "Nonparametric Identification and Estimation of Polychotomous Choice Models," *Journal of Econometrics*, 58.

McFadden, D. (1974): "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers of Econometrics*, ed. by P. Zarembka. Academic Press, New York.

Nevo, A. (2001): "Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica*, 69(2), 307–42.

Newey, W. K., and J. L. Powell (2003): "Instrumental Variable Estimation in Nonparametric Models," *Econometrica*, 71(5), 1565–1578.

Rivers, D. (1988): "Heterogeneity in Models of Electoral Choice," *American Journal of Political Science*, 32(3), 737–757.

Severini, T. A., and G. Tripathi (2006): "Some Identification Issues in Nonparametric Models with Endogenous Regressors," *Econometric Theory*.