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

October 23, 2008

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, endogenous choice characteristics, unknown heteroskedasticity, and correlated taste shocks. We consider full identification of the random utility model as well as identification of demand. 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 also show that key maintained hypotheses are testable.

*We have had helpful conversations on this topic with Hide Ichimura, Rosa Matzkin and Yuichi Kitamura. We also thank Sunyoung Park and seminar participants at Chicago, Harvard, Northwestern, M.I.T., Yale, and the Cowles Summer Conference on Identification of Economic Models 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, endogenous choice characteristics, unknown heteroskedasticity, and correlated taste shocks. We consider full identification of the random utility model as well as identification of demand. 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 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 choice characteristics and heterogeneous tastes in the spirit of Berry (1994), Berry, Levinsohn, and Pakes (1995), Nevo (2001), Petrin (2002), Berry, Levinsohn, and Pakes (2004) 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

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

or policy conclusions.

Our analysis demonstrates that with sufficiently rich data, random utility multinomial choice models featuring unobserved choice characteristics are identified without the parametric assumptions used in practice – typically, linear utility with independent additive or multiplicative taste shocks drawn from parametrically specified distributions. Our results may therefore lead to greater confidence in estimates and policy conclusions obtained using estimates of discrete choice demand models. In particular, they imply that parametric specifications used in practice can often be viewed as parsimonious approximations in finite samples rather than as 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 8). The first is the use of variation in observables to "trace out" the distribution of unobservables. Antecedents in the discrete choice literature include Manski (1985), Matzkin (1992), and Lewbel (2000). 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 contributions of observed and unobserved choice characteristics (see 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. Within-market variation in consumer attributes allows us to trace out a very flexible joint distribution of utilities while unobservable product characteristics are held fixed. Cross-market (and cross-product) variation in choice characteristics then allows identification of choice-specific unobservables. Full identification of the distribution of random utilities requires a "large support" condition, although a much less demanding condition on the support of observables is sufficient to identify demand.

In the following section we set up the choice framework 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. Here we provide results on identification of the joint distribution of preferences conditional on observed and unobserved choice We then move to discussion of several important extensions. In section 5 we show identifiability of demand under weaker conditions. Section 6 presents testable restrictions of key maintained hypotheses. In section 7 we suggest how our results can be extended to the case of data from a single market and a case in which only market level data are available. Having presented our results, we are then able to place our contributions within the context of the large literature on identification of multinomial choice models. After doing this in section 8, we conclude in section 9.

2 Model

2.1 Preferences and Choices

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 the researcher focuses on (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 $J_t = |\mathcal{J}_t| - 1$.[2] Each inside good

[2] 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

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). We will require at least one such measure for each $j \geq 1$ (we consider identification 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 represented by a random utility function

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

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 $\left\{\left(x_{jt}, \xi_{jt}, z_{ijt}\right)\right\}_{j \in \mathcal{J}_t}$, the random variation in the conditional indirect utilities is i.i.d. across individuals and markets.

markets in observable ways.

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 event ω_{it} from Ω. Obviously this places no restriction on the preferences of a given consumer, e.g., on the correlation of consumer-specific tastes for different goods or characteristics. This assumption also 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}.

Example 1. *A special case of the class of preferences we allow is generated by the linear random coefficients random utility model*

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

where, for example, $\beta_{it} = (\beta_{it}^{(1)}\left(\omega_{it}\right), \ldots, \beta_{it}^{(k)}\left(\omega_{it}\right))$ *is a vector of random coefficients and each* $\epsilon_{ijt}\left(\omega_{it}\right)$ *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* $\left(\beta_{it}^{(1)}, \ldots, \beta_{it}^{(k)}, \epsilon_{i1t}, \ldots, \epsilon_{iJ_t t}\right)$ *but requires that this joint distribution be the same for all* i, t, *and* $\left\{\left(x_{jt}, \xi_{jt}, z_{ijt}\right)\right\}_{j=1\ldots J}$.[3] *As an alternative, we could specify* $\beta_{it} = (\beta_{it}^{(1)}\left(z_{it}, \omega_{it}\right), \ldots, \beta_{it}^{(k)}\left(z_{it}, \omega_{it}\right))$

[3] This structure permits variation in J_t across markets. The realization of ω_{it} should be thought of as generating values of $\epsilon_{ijt} = \epsilon_j\left(\omega_{it}\right)$ for all possible choices j, not just those in the current choice set. Thus, the utility function defines preferences even over products not available. Note that here the joint distribution of $\left\{\epsilon_{ijt}\right\}_{j \in \mathcal{K}}$ 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.

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

$$u\left(x_{jt}, \xi_{jt}, z_{ijt}, \omega_{it}\right) > u\left(x_{kt}, \xi_{kt}, z_{ikt}, \omega_{it}\right) \qquad \forall k \in \mathcal{J}_t - \{j\}. \tag{2}$$

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

$$\tilde{u}_{ijt} = \phi_{it} z_{ijt}^{(1)} + \tilde{\mu}\left(x_{jt}, \xi_{jt}, z_{ijt}^{(2)}, \omega_{it}\right) \qquad \forall i, j = 1, \ldots, \mathcal{J}_t \tag{3}$$

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.[4] Here we have imposed two restrictions: (i) additive separability in a "vertical" component, $z_{ijt}^{(1)}$, of z_{ijt},[5] (ii) monotonicity in 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.[6] Because unobservables have no natural order,

---

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

[5] In the case of binary choice, additive separability is without loss when $z_{it}^{(1)}$ is a vertical characteristic. To see this, suppose the model is $y_{it} = 1\left\{u\left(x_t, \xi_t, z_{it}^{(1)}, z_{it}^{(2)}, \omega_{it}\right) > 0\right\}$. Since u is strictly increasing in $z_{it}^{(1)}$, this can be rewritten $y_{it} = 1\left\{z_{it}^{(1)} > u^{-1}\left(0; x_t, \xi_t, z_{it}^{(2)}, \omega_{it}\right)\right\}$. Letting $\mu\left(x_t, \xi_t, z_{it}^{(2)}, \omega_{it}\right) = -u^{-1}\left(0; x_t, \xi_t, z_{it}^{(2)}, \omega_{it}\right)$, we obtain a representation consistent with (3).

[6] 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 $\xi_{jt} \perp\!\!\!\perp c_{ijt}$.

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 (3) for each value of $\mathbf{z}_{it}^{(2)}$.[7]

2.2 Normalizations

Before discussing identification, we must have a unique representation of preferences for which the identification question can be posed. This requires several normalizations.

First, because unobservables enter non-separably and have no natural units we must normalize the location and scale of ξ_{jt}. For most of the paper we will assume without loss that ξ_{jt} 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] 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 Athey and Imbens (2007) point out. Their observation does not apply to our model in general. For example, conditional indirect utilities of the form $v_{ijt} = \xi_{jt} + z_{ijt}^{(1)}\beta_{it}$ are permitted by our model but do not lead to their their testable restriction. Nonetheless, 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:

$$u_{ijt} = z_{ijt}^{(1)} + \mu\left(x_{jt}, \xi_{jt}, z_{ijt}^{(2)}, \omega_{it}\right) \qquad \forall i, j = 1, \ldots, J^t. \tag{4}$$

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 (2) are determined by differences in utilities and we have not restricted correlation in the random components of utility across choices.

2.3 Examples

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 (4).

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}.

2.4 Observables and Structural Features of Interest

When we discuss the case of endogenous choice characteristics we will require excluded instruments, which we denote by $\tilde{\mathrm{w}}_{jt}$.[8] The observables then consist of $(y_{it}, \{x_{jt}, \tilde{\mathrm{w}}_{jt}, z_{ijt}\}_{j\in\mathcal{J}_t})_{i,t}$. To discuss identification, we treat their joint distribution as known. Loosely speaking, we consider the case of observations from a large number of markets, each with a large number of consumers.

The observables directly reveal the conditional choice probabilities

$$p_{ijt} = \Pr\left(y_{it} = j | \mathcal{J}_t, \{x_{kt}, \tilde{\mathrm{w}}_{kt}, z_{ikt}\}_{k\in\mathcal{J}_t}\right). \tag{5}$$

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.

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 $\{u_{ijt}\}_{j\in\mathcal{J}_t}$ conditional on any $\left(\mathcal{J}_t, \left\{x_{jt}, z_{ijt}, \xi_{jt}\right\}_{j\in\mathcal{J}_t}\right)$ in their support. These conditional distributions fully characterize the primitives of this model. We therefore refer to identification

[8] 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.

of these probability distributions as *full identification of the random utility model.*

We will also consider a type of partial identification: *identification of demand.* For many economic questions motivating estimation of discrete choice demand models, the joint distribution of utilities is not needed. For example, to discuss cross-price elasticities, equilibrium markups, or pricing/market shares under counterfactual ownership or cost structures, one requires identification of demand, not the full random utility structure. Identification of demand naturally requires less from the model and/or data than identification of the full probability distribution of preferences that underlie demand. In the multinomial choice setting, demand is fully characterized by the *structural choice probabilities*

$$\rho_j\left(\mathcal{J}_t, \{x_{jt}, \xi_{jt}, z_{ijt}\}_{j\in\mathcal{J}_t}\right) = \Pr\left(y_{it} = j | \mathcal{J}_t, \{x_{jt}, \xi_{jt}, z_{ijt}\}_{j\in\mathcal{J}_t}\right). \tag{6}$$

These conditional probabilities fully characterize demand. They are not directly observable from (5) because of the unobservables ξ_{jt}, which are typically correlated with at least some elements of x_{jt} (e.g., price).

3 Binary Choice with Exogenous Characteristics

Typically one will want to allow for endogeneity of at least one component of x_{jt}. In applications to demand estimation, price will typically be an observed characteristic that is correlated with the unobserved "quality" ξ_{jt} through the optimizing behavior of sellers.[9] However, we begin with the simple case of binary choice with exogenous x_{jt}. This will illustrate key elements of our approach and may be of independent interest, particularly given the prior attention to identification of binary choice models.

[9] 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.

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

Assumption 2 states that we consider here the special 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)).[10] 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 do use it. Here we show that Assumptions 1-3 are sufficient for full identification of the random utility model.

Begin by conditioning on a value of $\mathbf{z}_{it}^{(2)}$, which can then be suppressed. Rewrite (4) as

$$u_{it} = z_{it}^{(1)} + \mu_{it} \tag{7}$$

where we have let $\mu_{it} = \mu(x_t, \xi_t, \omega_{it})$ as shorthand. Holding the market 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

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

Assumption 3 guarantees that the distribution of $\mu_{it}|t$ (i.e., of $\mu(x_t, \xi_t, \omega_{it})$ within market t) is 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.

[10] 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.

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

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

With $F_{\mu_{it}|t}(\cdot)$ now known, each δ_t is known. Further, under Assumption 1, we can write

$$\delta_t = D\left(x_t, \xi_t\right) \tag{8}$$

for some function D that is strictly increasing in its second argument. Identification of each ξ_j then follows standard arguments. In particular, for $\tau \in (0,1)$ let $\delta^\tau(x_t)$ denote the τth quantile of $\delta_t|x_t$ across markets. By strict monotonicity of D in ξ_t, this quantile is unique. By (8) and the normalization of ξ_t

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

Since $\delta^\tau(x_t)$ is known 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 $F_{\mu_{it}|t}$. 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

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

for all $r \in \mathbb{R}$. With (7) this proves the following result.

Theorem 1. *Consider the binary choice setting with preferences given by (4). Under Assumptions 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 holding choice characteristics fixed. 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 typically be necessary in demand estimation. Our approach for the more general cases follows the same broad outline.

4 Multinomial Choice with Endogenous 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 prior work on multinomial choice by, e.g., 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 Assumption 2, we will require instruments. 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.

4.1 Identification with Fully Independent Instruments

Here 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. $\xi_{jt} \perp\!\!\!\perp (\mathrm{w}_{jt}, z_{ijt})\, \forall j,t$.

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}.[11] 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 $\varepsilon_{jt} \equiv \delta_{jt} - D\left(x_{jt},\tau\right)$, $E\left[B\left(x_{jt},\tau\right)\psi\left(x_{jt},\mathrm{w}_{jt},\tau\right)|\mathrm{w}_{jt}\right] = 0$ a.s. only if $B\left(x_{jt},\tau\right) = 0$ a.s., where $\psi(x,\mathrm{w},\tau) = \int_0^1 f_\varepsilon\left(\sigma B(x,\tau)|x,\mathrm{w}\right)\, d\sigma$.

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).[12] This condition 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 obtain the following result.

[11] Chernozhukov and Hansen's "rank invariance" property holds here because the same unobservable ξ_{jt} determines potential values of δ_{jt} for all possible values of the endogenous characteristics.

[12] They discuss sufficient conditions. We also consider an alternative to Assumption 6 below.

Theorem 2. *Under the representation of preferences in (4), 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 $\left(z_{i1t}^{(2)}, \ldots, z_{iJt}^{(2)}\right)$ and drop these arguments in what follows. Let $\mu_{ijt} = \mu\left(x_{jt}, \xi_{jt}, \omega_{it}\right)$ and observe that

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

Holding t fixed, $\mu_{ijt} \perp\!\!\!\perp z_{ijt}^{(1)}$ by Assumption 1. Assumption 4 then guarantees identification of the marginal distribution of each $\mu_{ijt}|t$. This implies identification of the conditional median $\delta_{jt} = med\left[\mu\left(x_{jt}, \xi_{jt}, \omega_{it}\right)|t\right]$. Since $med\left[\mu\left(x_{jt}, \xi_{jt}, \omega_{it}\right)|t\right] = med\left[\mu\left(x_{jt}, \xi_{jt}, \omega_{it}\right)|x_{jt}, \xi_{jt}\right]$, we can write

$$\delta_{jt} = D\left(x_{jt}, \xi_{jt}\right) \tag{9}$$

for some function D strictly increasing in ξ_{jt}. Under Assumptions 5 and 6, Theorem 4 of Chernozhukov and Hansen (2005) implies that D (and therefore each ξ_{jt}) is identified. Finally, observe that

$$\begin{aligned} p_{i0t} &= \Pr\left(z_{i1t}^{(1)} + \mu_{i1t} < 0, \ldots, z_{iJt}^{(1)} + \mu_{iJt} < 0\right) \\ &= \Pr\left(\mu_{i1t} < -z_{i1t}^{(1)}, \ldots, \mu_{iJt} < -z_{iJt}^{(1)}\right) \end{aligned} \tag{10}$$

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 $u_{ijt} = z_{ijt}^{(1)} + \mu_{ijt}$, the result follows. $\square$

4.2 Identification with Mean-Independent Instruments

Theorem 2 demonstrates the identifiability of a very general model of multinomial choice with endogeneity. A possible limitation is that Assumption 6 may be difficult to check and/or difficult to interpret. Whether there are useful sufficient conditions on economic 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 also enables us to relax the excludability restriction to require only mean independence.

Suppose each consumer i's conditional indirect utilities can be represented by

$$\tilde{u}_{ijt} = \beta_{it} z_{ijt}^{(1)} + \tilde{\mu}\left(x_{jt}, z_{ijt}^{(2)}, \omega_{it}\right) + \gamma_{it}\xi_{jt} \qquad j = 1, \ldots, \mathcal{J}_t \tag{11}$$

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 (3) 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 $z_{ijt}^{(1)}$ and ξ_{jt}, and with a nonparametric specification of $\tilde{\mu}$. A representation of preferences equivalent to (11) is

$$u_{ijt} = z_{ijt}^{(1)} + \mu\left(x_{jt}, \xi_{jt}, z_{ijt}^{(2)}, \omega_{it}\right) \qquad \forall i, j = 1, \ldots, \mathcal{J}_t \tag{12}$$

where now

$$\mu\left(x_{jt}, \xi_{jt}, z_{ijt}^{(2)}, \omega_{it}\right) = \frac{\tilde{\mu}\left(x_{jt}, z_{ijt}^{(2)}, \omega_{it}\right)}{\beta_{it}} + \frac{\gamma_{it}}{\beta_{it}}\xi_{jt}. \tag{13}$$

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

$$E\left[\frac{\gamma_{it}}{\beta_{it}}\right] = 1. \tag{14}$$

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 mean independence.

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

To show 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 (9). 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 (13) and the normalization (14) now we let

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

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, $E\left[B\left(x_{jt}\right)|\mathrm{w}_{jt}\right] = 0$ a.s. implies $B\left(x_{jt}\right) = 0$ a.s.

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

Theorem 3. *Under the utility representation (12), 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 (10). □

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 $x^{(1)}_{jt}$ to reveal $D(x_{jt})$ 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 $(\hat{x}^1, \ldots, \hat{x}^K) \times (\hat{\mathrm{w}}^1, \ldots, \hat{\mathrm{w}}^L)$, completeness corresponds to a full rank condition on the matrix $\{\sigma_{kl}\}$ where $\sigma_{kl} = \Pr(x_{jt} = \hat{x}^k | \mathrm{w}_{jt} = \hat{\mathrm{w}}^l)$.

5 Identification of Demand Using 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 $\left(z^{(1)}_{i1}, \ldots, z^{(1)}_{iJ_t}\right)$ 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 demand, i.e., 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 requires a single "common choice probability," i.e., at least one vector of choice probabilities that is attained in every market by conditioning on appropriate vectors $\left(z^{(1)}_{i1t}, \ldots, z^{(1)}_{iJ_tt}\right)$ for each market t. By contrast, the large support condition requires every point in the simplex to be a common choice probability.

We also 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 these results under a somewhat more restrictive specification of preferences. Up to this qualification, however, these results should be a comforting. Demand is identified under much weaker support conditions. And although we show full identifiability of the random utility model only with the large support assumption, the result is not knife-edge: the tails of the large support are needed only to determine the tails of the joint distributions of utilities themselves.

5.1 Binary Choice

As before, we begin with binary choice to illustrate our main insights. We first consider the general specification of preferences in (4) above. We then consider the a more restrictive specification (analogous to (12)) that appears to be more useful for the multinomial case. For both cases 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.[13] This is a much weaker requirement and is likely to be satisfied in many applications.

5.1.1 General Case

Consider the specification of preferences in (4). In the binary choice case, the consumer chooses the inside good if (fixing $z_{it}^{(2)}$ and suppressing it)

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

[13] 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.

With Assumption 9, for each market t we can identify the value z_{it}^{τ} such that

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

Observe that 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

$$z_t^{\tau}=\zeta\left(x_t,\xi_t;\tau\right) \tag{15}$$

for some function $\zeta(\cdot;\tau)$ that is strictly decreasing in ξ_t. This strict monotonicity is the key idea here: holding x_t fixed, markets with high values of z_t^{τ} are those with low values of the unobservable ξ_t.

Identification of the function $\zeta(\cdot;\tau)$ and, therefore, of each ξ_t, then follows from (15) 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 the structural choice probabilities

$$\rho\left(x_t,\xi_t,z_{it}\right)=\Pr\left(y_{it}=1|x_t,\xi_t,z_{it}\right) \tag{16}$$

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

Theorem 4. *In the binary choice model with preferences given by (4), 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.*

We require only one common choice probability. If there is more than one, each provides additional information about the distribution of $u_{i1t}|x_t,z_{it},\xi_t$. In particular, we can identify a $\zeta(\cdot;\tau)$ 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 demonstrates the notion of "continuity" described 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.

5.1.2 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 section 4.2. Here we also impose a further restriction on the utility specification in (11)—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

$$u_{ijt} = z_{ijt}^{(1)} + \mu\left(x_{jt}, z_{ijt}^{(2)}, \omega_{it}\right) + \xi_{jt} \qquad \forall i, j = 1, \ldots, \mathcal{J}_t. \tag{17}$$

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 (17) is still a generalization relative to the literature, even without our relaxation of the the large support assumption here.

Theorem 5. *In the binary choice model with preferences given by (17), 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 $\mathbf{z}_{it}^{(2)}$ and suppressing it, we have

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

by definition of z_t^{τ}. 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

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

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

$$z_t^\tau = \zeta\left(x_t; \tau\right) - \xi_t. \tag{18}$$

For any τ, (18) defines a nonparametric regression equation. Identification of $\zeta\left(\cdot; \tau\right)$ (and therefore ξ_t) then follows from the results of Newey and Powell (2003) as before. As in the preceding section, this is sufficient to determine the structural choice probabilities $\rho\left(x_t, \xi_t, z_{it}\right) = \Pr\left(y_{it} = 1|x_t, \xi_t, z_{it}\right)$ at all points $\left(x_t, \xi_t, z_{it}\right)$ of support. $\square$

Note that 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 Multinomial Choice

For the multinomial case we will maintain the more restrictive representation of preferences in (17), where

$$u_{ijt} = z_{ijt}^{(1)} + \mu\left(x_{jt}, z_{ijt}^{(2)}, \omega_{it}\right) + \xi_{jt} \qquad \forall i, j = 1, \ldots, \mathcal{J}_t. \tag{19}$$

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 $\mathbf{z}_t^q = (z_{1t}^q, \ldots, z_{1t}^q) \in \mathrm{supp}\left(z_{i1t}^{(1)}, \ldots, z_{iJ_t t}^{(1)}\right)$ such that $q_j = \Pr(y_{it} = j\ |x_{1t}, \ldots, x_{J_t t}, z_{i1t}, \ldots, z_{iJ_t t})_{\mathbf{z}_{it}^{(1)} = \mathbf{z}_t^q}$ 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).[14] 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.

Theorem 6. *In the multinomial choice model with preferences given by (19), 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 (19), choice probabilities depend on the sums

$$\lambda_{ijt} \equiv z_{ijt}^{(1)} + \xi_{jt} \tag{20}$$

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

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

From (20) and Assumption 10, for all $(x_{1t}, \ldots, x_{J_t t})$ there is a unique vector

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

[14] See Gandhi (2008) for an extension.

such that

$$\lambda_j(\mathbf{x}_t, q) = \xi_{jt} + z_{jt}^q \tag{21}$$

and

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

From (21),

$$z_{jt}^q = \lambda_j(\mathbf{x}_t, q) - \xi_{jt} \quad \forall j, t. \tag{22}$$

These equations identify the functions $\lambda_j(\cdot, q)$ 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

The models we have considered are quite general but rely on two important assumptions: (i) existence of a vertical additively separable observable, $z_{ijt}^{(1)}$; (ii) a scalar vertical choice-specific unobservable, ξ_{jt}. Here we show that both assumptions imply testable restrictions.[15]

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 7. *Suppose preferences are characterized by (4). Then under Assumption 1,* $p_{ijt} \equiv \Pr(y_{it} = j | \mathcal{J}_t, \{x_{jt}, w_{jt}, z_{ikt}\}_{k \in \mathcal{J}_t})$ *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

[15] 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)).

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

$$\rho_j\left(\mathcal{J}_t, \{x_{jt}, \xi_{jt}, z_{ijt}\}_{j\in\mathcal{J}_t}\right) = \int_{A_j} dF_u\left(u_1, \ldots, u_{J_t} | \{x_{jt}, \xi_{jt}, z_{ijt}\}_{j\in\mathcal{J}_t}\right) \quad \forall j. \tag{23}$$

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 (23). 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 (23) (but did rely on the linearity of utilities in $z_{ijt}^{(1)}$), we have the following.

Theorem 8. *Under the hypotheses of Theorem 2 or Theorem 3, the overidentifying restrictions* (23) *must hold.*

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 4 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 9. *In the binary choice model with preferences given by (4), 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 $\xi_t\left(z_t^\tau; \tau, x_t\right) > 0$.*[16] *Recall that each z_t^τ is observable and defined such*

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

that $\Pr(\nu_{it}\xi_t < z_t^\tau) = \tau$*. Letting* Φ *denote the standard normal CDF, this requires*

$$\Phi\left(\frac{z_t^\tau}{\xi_t}\right) = \tau \quad \forall t. \tag{24}$$

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, violating the restriction from Theorem 9.*

Theorem 10. *In the binary choice model with preferences given by (4), 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* $\xi_t(z_t^\tau; \tau, x_t) = \xi_t\left(z_t^{\tau'}; \tau', x_t\right)$ *for all* t*.*

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

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* $\nu_{it} > 1-\tau$ *choose the inside good. Thus, the value of* z_t^τ *is determined by the preferences of the consumer with* $\nu_{it} = 1-\tau$*. Now consider the* $\xi_t(\tau)$ *inferred under the incorrect assumption of a scalar unobservable. From the observations above, when* $\tau > 1/2$ *we have* $\xi_t(\tau) = F_{\xi^1+\xi^2}\left(\xi_t^1 + \xi_t^2\right)$ *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*. Similarly, 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 Extensions

7.1 A Single Market

Here we depart from the rest of the paper by imagining a different sampling structure. Suppose there is a single choice set (market). Obviously the intuition of using variation across markets to identify the choice-specific unobservables cannot apply. However, there is still information about product unobservables coming from variation across products within the market. With a single market, identification can be obtained from this within-market variation if one is willing to consider the case of a large (infinitely large) choice set.

We briefly discuss this case here for two reasons. One is that this links to previous work (e.g., Berry, Linton, and Pakes (2004)) relying on "large J" asymptotic approximations for parametric econometric models of discrete choice. The other is that this may provide a useful benchmark for understanding identification in applications with little or no variation in choice sets: in principle, a single choice set can suffice.

Suppressing $z_{ij}^{(2)}$ and dropping the superfluous market subscript t, the model in (4) is now

$$u_{ij} = z_{ij}^{(1)} + \mu\left(x_j, \xi_j, \omega_{it}\right) \qquad \forall i, j \in \mathcal{J}.$$

For any choice set $\mathcal{J}$ with $|\mathcal{J}| = J + 1$, consider the choice probability for the outside good:

$$p_0 = \Pr\left(z_{i1}^{(1)} + \mu\left(x_1, \xi_1, \omega_{it}\right) < 0, \ldots, z_{iJ}^{(1)} + \mu\left(x_J, \xi_J, \omega_{it}\right) < 0\right).$$

The large support condition (Assumption 3) implies identification of the joint distribution of $\left(\mu\left(x_1, \xi_1, \omega_{it}\right), \ldots, \mu\left(x_J, \xi_J, \omega_{it}\right)\right)$. One implication is that each

$$\delta_j \equiv med\ \mu\left(x_j, \xi_j, \omega_{it}\right) | x_j, \xi_j$$

can be considered known, although the values of ξ_j conditioned on are still unknown. However, since we can write

$$\delta_j = D\left(x_j, \xi_j\right) \tag{25}$$

identification of each ξ_j is equivalent to identification of the regression model (25). Because D strictly increasing in ξ_j, as $J \to \infty$ the identification result of Chernozhukov and Hansen (2005) delivers identification of the function D and, thus, each ξ_j. This then provides identification of the joint distribution of utilities $(u_{i1}, \ldots, u_{iJ})$ given $(x_1, \ldots, x_J, \xi_1, \ldots, \xi_J)$. Of course, with $J \to \infty$, this determines the joint distribution of utilities for all subsets of choices—i.e., for all possible choice sets.

7.2 Market Level Data

In many applications one is forced to work without micro data linking choices to individual characteristics, relying instead on market level choice probabilities (i.e., market shares). Berry and Haile (2008) explore identification for such settings using a different approach from the "two-step" identification arguments we have used above. However, there is at least one case in which the ideas in the present paper can be directly applied to the case of market level data.

Eliminating the micro data z_{ijt} from the model, the observables are now (y_{it}, x_{jt}). Note that each x_{jt} could contain attributes of products j or attributes of markets t. Partition x_{jt} into $\left(x_{jt}^{(i)}, x_{jt}^{(ii)}\right)$ and suppose preferences can be represented by conditional indirect utilities of the form

$$u_{ijt} = x_{jt}^{(i)} + \mu(x_{jt}^{(ii)}, \xi_{jt}, \omega_{it}). \tag{26}$$

Assume that the set of markets can be partitioned into *market groups* Γ such that for all $t \in \Gamma$, $\left(x_{jt}^{(ii)}, \xi_{jt}\right) = \left(x_{j\Gamma}^{(ii)}, \xi_{j\Gamma}\right)$. One natural example of such an environment is that of a national industry (e.g., the U.S. automobile industry) in which the physical products themselves are identical across regions of the nation, but regions may differ in average income, product prices (e.g., due to f.o.b. pricing), prices of complementary goods (e.g., gasoline), availability of substitute goods (e.g., public transportation), etc.

For simplicity, we illustrate the argument formally only for the case of full identification with exogenous product characteristics. However, it will be clear that all the identification results obtained above have analogs in this setting.

Assumption 11. supp $(x^{(i)}_{1t}, \ldots, x^{(i)}_{J_t t}) | (x^{(ii)}_{1t}, \ldots, x^{(ii)}_{J_t t}) = \mathbb{R}^{J_t}\ \forall t.$

Assumption 11 is different from the parallel large support Assumption 4 in requiring sufficient variation in an exogenous product characteristic rather than an individual-product observable. The role of this assumption is the same, however: to trace out the distribution of the random component of (26) within each market group.

Under these assumptions, the setup is isomorphic to that in section 4. Variation in $x^{(i)}_{jt}$ across market groups at the limit $x^{(i)}_{j't} \to -\infty\ \forall j' \neq j$ identifies the distribution of $\mu_i\left(x^{(ii)}_{j\Gamma}, \xi_{j\Gamma}\right)$ exactly as in section 4. Letting $\delta\left(x^{(ii)}_{j\Gamma}, \xi_{j\Gamma}\right) = E\left[\mu_i\left(x^{(ii)}_{j\Gamma}, \xi_{j\Gamma}\right) | x^{(ii)}_{j\Gamma}, \xi_{j\Gamma}\right]$, identification of the function $\delta\left(x^{(ii)}_{j\Gamma}, \xi_{j\Gamma}\right)$ (and therefore each $\xi_{j\Gamma}$) follows exactly as in the previous sections. With each $\xi_{j\Gamma}$ and the distribution of $\mu_i\left(x^{(ii)}_{j\Gamma}, \xi_{j\Gamma}\right)$ known, the joint distribution of $\{u_{ijt}\}_{j\in\mathcal{J}_t}$ is uniquely determined at any $\left(\mathcal{J}_t, \left\{(x_{jt}, \xi_{jt})\right\}_{j\in\mathcal{J}_t}\right)$ in their support.

Because the setup here is isomorphic to that for the case of micro data, the extensions to the case of endogenous characteristics (elements of $x^{(ii)}_{jt}$), a separable error structure, and identification of demand with limited support follow directly as well.

8 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 setup 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. In addition, our attention to identification of demand as an alternative to full identification of the random utility model appears to be new to the literature.

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. Fox and Gandhi (2008) explores identifiability of several related models, including a flexible model of multinomial choice in which consumer types are multinomial and utility functions are determined by a finite parameter vector. They suggest that our approach for allowing choice-specific unobservables and endogenous choice characteristics could be adapted to their framework. 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 (e.g., Blundell and Powell (2004)).

9 Discussion

One reason we have been able to make progress in well worn territory is our recognition that, aside from welfare analysis, motivations for demand estimation in practice often require identification only of demand, not the full random utility model. To identify demand the critical step is uncovering choice-specific unobservables ξ_{jt} and we have shown that these are identified with under relatively weak conditions. Another source of our progress is the generality of the model we consider. This may be counterintuitive; however, focusing on the nonparametric joint distribution of utilities *directly* rather than on distributions of parameters in a more restrictive specification naturally limits the dimensionality of the primitives of interest (a J_t dimensional joint distribution of conditional utilities) to the dimension of the observables (a J_t-dimensional joint distribution of conditional market shares).

The generality of our model of preferences comes with some costs, however. One is that some out-of-sample counterfactuals will not be identifiable.[17] 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 limitation concerns welfare. Our specification of preferences (4) incorporates quasilinearity preferences and can therefore be used to characterize changes in utilitarian social welfare (in aggregate, or across subpopulations defined by observables).[18] However, (4) lacks the structure required for welfare analysis that depends on the *distribution* of welfare changes. Characterization of Pareto improvements, for example, would 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

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

[18] 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 (4). 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.

given individual. This points out a limitation of nonparametric random utility models as a theoretical foundation for some kinds of welfare analysis; in practice, such welfare calculations will require additional *a priori* structure.

An example of a model with sufficient structure to address all welfare questions is the linear random coefficients random utility model

$$u_{ijt} = x_{jt}\beta_{it} + z_{ijt}\gamma + \xi_{jt} + \epsilon_{ijt}. \tag{27}$$

This generates a special case of the general model of preferences we have considered in this paper. We provided conditions for identification of the joint distribution of $\left(\{u_{ijt}\}_j \mid \left\{x_{jt}, \xi_{jt}, z_{ijt}\right\}_j\right)$ for all $\left\{x_{jt}, \xi_{jt}, z_{ijt}\right\}_j$ in their support. Going from here to identification of the model in (27) is then equivalent to identification of a linear random coefficients regression model. Beran and Hall (1992) and Beran, Feuerverger, and Hall (1996) have discussed sufficient conditions, which involve regularity and support requirements we did not need for our results.

Finally, we point out that an important distinction between our work and much of the prior literature is our neglect of estimation. Although our identification proofs may suggest new nonparametric or semiparametric estimation approaches, additional work would be needed. However, our main objective has been to explore identification in a very general random utility discrete choice 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 Economy*, 115(5), 588–638.

Beran, R., A. Feuerverger, and P. Hall (1996): "On Nonparametric Estimation of Intercept and Slope Distributions in Random Coefficient Regression," *The Annals of Statistics*, 240, 2569–2592.

Beran, R., and P. Hall (1992): "Estimating Coefficient Distributions in Random Coefficient Regressions," *The Annals of Statistics*, 20, 1970–1984.

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., O. Linton, and A. Pakes (2004): "Limit Theorems for Differentiated Product Demand Systems," *Review of Economic Studies*, 71(3), 613–614.

Berry, S. T., and P. A. Haile (2008): "Identification of Discrete Choice Demand 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.

Gandhi, A. (2008): "On the Nonparametric Foundations of Discrete Choice Demand Estimation," Discussion paper, University of Wisconsin-Madison.

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.

Petrin, A. (2002): "Quantifying the Benefits of New Products: The Case of the Minivan," *JPE*, 110(4), 705–729.

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*.