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

February 2, 2009

Abstract

We consider identification of nonparametric random utility models of multinomial choice using "micro data," i.e., observation of the characteristics and choices of individual consumers. Our model of preferences nests random coefficients discrete choice models widely used in practice with parametric functional form and distributional assumptions. However, the model is nonparametric and distribution free. It allows choice-specific unobservables, endogenous choice characteristics, unknown heteroskedasticity, and high-dimensional correlated taste shocks. Under standard "large support" and instrumental variables assumptions, we show identifiability of the random utility model. We demonstrate robustness of these results to relaxation of the large support condition and show that when it is replaced with a weaker "common choice probability" condition, the demand structure is still identified. We 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 received useful comments from Sunyoung Park and participants in several conferences and seminars.

1 Introduction

We consider identification of nonparametric random utility models of multinomial choice using "micro data," i.e., observation of the characteristics and choices of individual consumers.[1] Our model of preferences nests random coefficients discrete choice models widely used in practice with parametric functional form and distributional assumptions. However, the model is nonparametric and distribution free. It allows choice-specific unobservables, endogenous choice characteristics, unknown heteroskedasticity, and high-dimensional correlated taste shocks. Under standard "large support" and instrumental variables assumptions, we show identifiability of the random utility model, i.e., of (i) the choice-specific unobservables and (ii) 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 it is replaced with a 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), Chintagunta, Jain, and Vilcassim (1991)), urban economics (e.g., Bayer, Ferreira, and McMillan (2007)), education (e.g., Hastings, Staiger, and Kane (2007)), migration (e.g., Kennan and Walker (2006)), voting (e.g., Rivers (1988)), and health economics (e.g., Ho (2007)). We focus in particular on discrete choice random utility models allowing both unobserved choice characteristics and heterogeneous tastes, in the spirit of Berry (1994), Berry, Levinsohn and Pakes (1995, 2004), Nevo (2001), Petrin (2002), and a large related literature. Although this class of models has been applied to research in many areas, the sources of identification have not been fully

[1] In Berry and Haile (2008b) we consider identification using market level data, where one observed only conditional choice probabilities (market shares).

understood. Without such an understanding it is difficult to know what qualifications are necessary when interpreting estimates or policy conclusions.

Our analysis demonstrates that with sufficiently rich micro data, random utility multinomial choice models featuring unobserved choice characteristics are identified without the parametric or distributional assumptions used in practice—typically, linear utility with independent additive and/or multiplicative taste shocks drawn from parametrically specified distributions. Our results may therefore lead to greater confidence in estimates and policy conclusions obtained in empirical work based on discrete choice models. In particular, parametric specifications used in estimation 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.

An important strategy in our work is modeling utility as a fully general nonparametric random function of observed and unobserved characteristics. This contrasts with the usual approach of building up randomness from random coefficients and/or other taste shocks. In addition to enabling us to consider a very general random utility model, this formulation leads us to focus directly on identification of the conditional joint distribution of utilities rather than the joint distribution of random coefficients and/or other taste shocks. The advantage of this might be unexpected: a natural intuition is that added structure on the way randomness enters would aid identification. However, whereas the conditional distribution of utilities has the same dimension as the observable conditional choice probabilities (i.e., the dimension of the choice set), most random coefficients models involve taste shocks of dimension strictly larger than that of the choice set. Focusing directly on the joint distribution of utilities naturally leads to primitives with the "correct" dimension without strong distributional or functional form restrictions.

A second key aspect of our work is our explicit modeling of choice-specific unobservables. Although this is standard in the applied literature, much of the prior work on identification of discrete choice models has embedded the sources of randomness in preferences and the sources of endogeneity in the same random variables. In applications to demand estimation, an endogeneity problem typically arises because some observed choice characteristics (price being a leading example) depend on unobserved choice characteristics. For such environments, explicitly modeling choice-specific unobservables enables one to define counterfactuals involving changes in endogenous characteristics within a model of heteroskedastic random utilities. For example, our formulation allows us to characterize demand elasticities, which require evaluating the effects of a change in price (including resulting changes in the variance or other moments of random utilities), holding unobserved product characteristics fixed.

A third novel component of our work is its exploration of both identification of the full model and identification of "demand"—i.e., the mapping from observed and unobserved characteristics to the vector of choice probabilities. For many questions motivating estimation of discrete choice models, knowledge of this demand structure suffices. Not surprisingly, identification of demand can be obtained under weaker conditions than those giving full identification of the random utility model.

Despite these differences from the prior literature, we rely heavily on several well known ideas. One 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, 1993), Lewbel (2000), Honoré and Lewbel (2002), and Briesch, Chintagunta, and Matzkin (2005), among others. A second idea is the use of exogenous variation in choice sets to decompose variation in the distribution of utilities into the contributions of observed and unobserved characteristics. This strategy has been exploited in parametric discrete choice models by, e.g., Berry (1994) and Berry, Levinsohn and Pakes (1995, 2004). Here we rely heavily on results from the recent literature on nonparametric identification of regression models using instrumental variables, particularly Newey and Powell (2003) and Chernozhukov and Hansen (2005).

In the following section we provide some additional discussion of related literature. We then set up the choice framework and define the observables and structural features of interest in section 3. Section 4 provides an illustration of key lines of argument in a simple case: binary choice with exogenous characteristics. Section 5 addresses full identification in the case of multinomial choice with endogeneity. There we consider two alternative instrumental variables conditions that deliver full identification of the model. In section 6 we show identifiability of demand under weaker conditions and provide an observation about robustness of the full identification results to relaxation of support conditions. Section 7 discusses testable restrictions of key maintained hypotheses. In section 8 we show how our results can be extended to one type of environment in which only market level data are available. We conclude in section 9.

2 Literature

There is a large literature on identification of discrete choice models and we cannot attempt a complete review here. However, important early work on identification of discrete choice models includes Manski (1985, 1987, 1988) and Matzkin (1992, 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 an additively separable taste shock for each choice. Extensions to nonadditive models can be found in Matzkin (2005, 2007a, 2007b, 2008).

Identification of heterogeneous preferences has 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 generalizations of the linear random coefficients model. All of these consider models of utility that is linear in at least one characteristic, which we will also require. We add to this literature by allowing a more general representation of preferences and providing a more complete treatment of choice specific unobservables and endogeneity.[2]

Lewbel (2000) considers identification in a semiparametric model with an additive special regressor (which we also have in much of this paper), and allowing for endogeneity or heteroskedasticity. In particular, he allows for an additive stochastic component whose distribution can vary with observables. Our model of preferences differs from his in two main dimensions. First, we relax functional form restrictions; for example, we do not require mean effects of observables to enter separably from unobservables. Second, we make a distinction between choice-specific unobservables and individual heterogeneity in preferences. This enables us to identify important primitive features, including demand elasticities, but also rules out the possibility that choice sets depend on unobserved consumer tastes. Our model is appropriate for most applications to demand estimation, but inappropriate evaluating the treatment effects on discrete outcomes in environments with selection on individual unobservables.

Honoré and Lewbel (2002) consider a binary panel version of the model in Lewbel (2000), relying on linearity of the composite error term and focusing on identification of a slope parameter. Altonji and Matzkin (2005) consider a similar but nonparametric model. For discrete choice models, their results focus on identification and estimation of local average responses. Other work considering models similar to that Lewbel (2000) includes Hong and Tamer (2004), Blundell and Powell (2004), Lewbel (2005), and Magnac and Maurin (2007). All consider linear semiparametric models, most limiting attention to binary outcomes.

Matzkin (2004) (section 5.1) considers a model incorporating choice-specific unobservables and an additive preference shock, but in a model without random coefficients or other sources of heteroskedasticity.[3] Hoderlein (2008) allows for both heteroskedasticity and endogeneity in the case of binary choice, focusing on identification of an average derivative. Blundell and Powell (2004), Matzkin (2004), and Hoderlein (2008) limit attention binary

[2] Concurrent work by 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 incorporating choice-specific unobservables and endogenous choice characteristics could be adapted to their framework.

[3] See also Matzkin (2007a, 2007b).

choice in semiparametric triangular models, leading to the applicability of control function methods or the related idea of "unobserved instruments."[4] For binary choice demand, triangular models can be appropriate when price depends only on the unobserved demand shock, but not on a cost shock as well. In the case of multinomial choice, standard supply models for differentiated products markets imply that each price depends on the entire vector of demand shocks (and cost shocks, if any). This appears to preclude the use of control function methods except in binary choice applications where there are no supply shocks.[5]

Finally, we must 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, significant additional work would be needed.

3 Model

3.1 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 are defined to be 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),

[4] See also Lewbel (2000), Honoré and Lewbel (2002), Altonji and Matzkin (2005), and Petrin and Train (2009).

[5] However, for the case of market level data, Berry and Haile (2008b) uses a related approach of inverting a multivariate supply and demand system to recover the entire vector of shocks to supply and demand.

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 $J_t = |\mathcal{J}_t| - 1$.[6] Each inside good j has observable (to us) characteristics x_{jt}. Among other things, x_{jt} can include product dummies and price. Unobserved choice characteristics are characterized by an index ξ_{jt}, which may also vary across markets; indeed, ξ_{jt} may also reflect the unobserved taste for choice j in market t. 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 both consumer- and choice-specific. This arises, for example, when there are interactions between consumer demographics and product characteristics (say, family size and automobile size), or from 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 $j \geq 1$. This is standard in applications that specify consumer preferences as determined in part by consumer characteristics.[7] Let $\mathbf{x}_t = (x_{1t}, \ldots, x_{J_t t})$ and $\mathbf{z}_{it} = (z_{i1t}, \ldots, z_{iJ_t t})$. Let χ denote the support of $\left(x_{jt}, \xi_{jt}, z_{ijt}\right)$.

We consider preferences represented by a random utility model. Each consumer i in market t has a conditional indirect utility function $u_{it} : \chi \to \mathbb{R}$. However, consumers have heterogeneous tastes, even conditional on observables. Thus, from the perspective of the researcher, each utility function u_{it} can be viewed as a random draw from a set $\mathcal{U}$ of permissible functions $\{u : \chi \to \mathbb{R}\}$ (we will discuss our assumptions on the set $\mathcal{U}$ below). Formally, we define the random function u_{it} on χ as follows. Let $(\Omega, \mathcal{F}, \mathbb{P})$ denote a probability space.

[6] 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.

[7] Without this—for example if consumer characteristics affect only tastes for the outside good—the identification problem is identical to that in the case of market-levl data (see Berry and Haile (2008b)), conditional any consumer characteristics.

Given any $\left(x_{jt}, \xi_{jt}, z_{ijt}\right) \in \chi$,

$$u_{it}\left(x_{jt}, \xi_{jt}, z_{ijt}\right) = u\left(x_{jt}, \xi_{jt}, z_{ijt}, \omega_{it}\right) \tag{1}$$

where $\omega_{it} \in \Omega$, and u is measurable in ω_{it}.[8] Without loss, the the draw from the sample space Ω determining the realized function u_{it} is specified as independent of the arguments of the function, $\left(x_{jt}, p_{jt}, \xi_{jt}\right)$.[9] We add to this an assumption of menu-independent preferences:

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

Relative to the standard formulation of a random function on χ, this assumption merely rules out the possibility that preferences over a given set of good depends on the other goods available.[10]

Aside from this menu-independence and the restriction to scalar choice-specific unobservables, our representation of preferences is so far fully general. For example, it allows arbitrary correlation of consumer-specific tastes for different goods or characteristics, as well as arbitrary heteroskedasticity across products and/or across consumers with different z_{ijt}. The following example illustrates how our model can be specialized to a more familiar semi-parametric models.

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

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

[8] Despite the similarity of our notation to that in Matzkin (2007b) and Matzkin (2007a), in our formulation ω_{it} is not a random variable (or random vector) but an elementary event in the sample space Ω. Each ω_{it} maps to a different utility function. As the examples below illustrate, in special cases of our model the realization of ω_{it} could determine the realizations of a large number of random variables with arbitrary joint distribution.

[9] Arbitrary dependence of the distribution of u_{ijt} on $\left(z_{ijt,}x_{jt}, \xi_{jt}\right)$ is permitted through the function u.

[10] See Block and Marschak (1960), Falmagne (1978), and Barbera and Pattanaik (1986) for testable restrictions that follow from this assumption.

Here the random variables $\left(\beta_{it}^{(1)},\ldots,\beta_{it}^{(K)},\epsilon_{i1t},\ldots,\epsilon_{iJ_tt}\right)$ *can be defined on the probability space* $(\Omega,\mathcal{F},\mathbb{P})$*, for example as* $\beta_{it}^{(k)}=\beta_{it}^{(k)}(\omega_{it})$ *and* $\epsilon_{ijt}=\epsilon_j(\omega_{it})$*. With this specification, Assumption 1 allows an arbitrary joint distribution of* $\left(\beta_{it}^{(1)},\ldots,\beta_{it}^{(K)},\epsilon_{i1t},\ldots,\epsilon_{iJ_tt}\right)$ *but requires that it be the same for all* i*,* t*, and* $\left\{\left(x_{jt},\xi_{jt},z_{ijt}\right)\right\}_{j=1\ldots J}$*.*[11] *This specification of* $\left(\beta_{it},\{\epsilon_{ijt}\}_j\right)$ *is both more general than typically allowed in the literature and more restrictive than required by our framework, even within a linear random coefficients model. Also permitted would be* $\beta_{it}=(\beta_{it}^{(1)}(z_{it},\omega_{it}),\ldots,\beta_{it}^{(K)}(z_{it},\omega_{it}))$ *and* $\epsilon_{ijt}=\epsilon\left(x_{jt},\xi_{jt},\omega_{it}\right)$*, where* z_{it} *is a vector of individual characteristics that do not vary across* j*. Indeed, we could generalize further by specifying* $\epsilon_{ijt}=\epsilon\left(x_{jt},\xi_{jt},z_{ijt},\omega_{it}\right)$*; however, then the sum* $x_{jt}\beta_{it}+z_{ijt}\gamma+\xi_{jt}$ *in (2) would be redundant and the model would collapse to our completely general specification (1).*

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

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

For much of the paper we will rely on a separability restriction on preferences. 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

[11] This structure permits variation in J_t across markets. In our formulation, the realization of ω_{it} determines a consumer's utility function. Thus the realization of ω_{it} should be thought of as generating values of the random variables $\epsilon_{ijt}=\epsilon_j(\omega_{it})$ for all possible choices j, not just those in the current choice set. Note that under Assumption 1 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.

preferences with the form

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

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.[12]

Here we have imposed two restrictions:

(i) additive separability in a "vertical" attribute $z_{ijt}^{(1)}$

(ii) monotonicity in ξ_{jt}.

An important implication of separability is that $z_{ijt}^{(1)}$ is independent of the stochastic component of $u_{it}\left(x_{jt}, \xi_{jt}, z_{ijt}\right)$, conditional on $x_{jt}, \xi_{jt}, z_{ijt}^{(2)}$. This gives $z_{ijt}^{(1)}$ the form of a "special regressor" as in Lewbel (2000). In particular, we rely on the separability restriction to provide a mapping between units of (latent) utility and units of (observable) choice probabilities.[13] Below we will consider cases with and without a large support assumption on $z_{ijt}^{(1)}$. 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 allow heterogeneity in tastes for ξ_{jt} and for the vertical observable $z_{ijt}^{(1)}$. Furthermore, we allow a different representation (4) for each value of $\mathbf{z}_{it}^{(2)}$.[14] We show in section 7 that both (i) and (ii) have testable implications.

---

[12] 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.

[13] We can allow $z_{ijt}^{(1)}$ to be an index $g(c_{ijt})$ where c_{ijt} is a vector, as long as $\xi_{jt} \perp\!\!\!\perp c_{ijt}$. Within a market (so that all x_{jt} and ξ_{jt} are fixed) utilities have the form $\tilde{u}_{ijt} = g(c_{ijt}) + \mu_{ijt}$ with $\xi_{jt} \perp\!\!\!\perp \mu_{ijt}$. If $g(\cdot)$ is linear, identification of $g(\cdot)$ follows by standard results (e.g., Manski (1985)). Identification of nonlinear $g(\cdot)$ can be obtained under restrictions considered in Matzkin (1993).

[14] 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.

3.2 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

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

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.

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.

3.3 Observables and Structural Features of Interest

For most results we will require excluded instruments, which we denote by $\tilde{\mathrm{w}}_{jt}$.[15] The observables then consist of

$$\left(y_{it}, t, \{x_{jt}, \tilde{\mathrm{w}}_{jt}, z_{ijt}\}_{j\in\mathcal{J}_t}\right).$$

To discuss identification, we treat their joint distribution as known. In particular, we take the conditional probabilities

$$p_{ijt} = \Pr_{\mathbb{P}}\left(y_{it} = j | t, \{x_{kt}, \tilde{\mathrm{w}}_{kt}, z_{ikt}\}_{k\in\mathcal{J}_t}\right) \tag{6}$$

as known. Loosely speaking, we consider the case of observations from a large number of markets, each with a large number of consumers. Thus, interpreting this as a panel setting, we consider observations for "large T and large N."

We write $\Pr_{\mathbb{P}}$ in (6) to make clear that we do not permit selection on consumer-specific unobservables. Thus, the choice probabilities we observe are those of the full population of consumers—i.e., for the population of utility functions defined by draws from Ω governed by $\mathbb{P}$. This rules out applications where some components of $(\mathcal{J}_t, \{x_{kt}, z_{ikt}\}_{k\in\mathcal{J}_t})$ are chosen in response to consumer unobservables.[16] Our assumption, which treats ξ_{jt} as the unobservable responsible for the endogeneity problem, is appropriate for most applications to multinomial choice, where the same choices are offered to all consumers in a market.

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 $\left\{\xi_{jt}\right\}_{j\in\mathcal{J}_t}$ and 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

[15] 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.

[16] As an illustration, in the linear random coefficients model of Example 1, we permit arbitrary correlation between $(\beta_i, \epsilon_{ijt})$ and (z_{ijt}, x_{jt}) but view this as a structural feature in the population, not the result of selection of (z_{ijt}, x_{jt}) on the prefeence shocks $(\beta_i, \epsilon_{ijt})$.

support. These conditional distributions fully characterize the primitives of this model. We therefore refer to identification 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 distribution of preferences. In the multinomial choice setting, demand is fully characterized by the *structural choice probabilities*

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

These conditional probabilities 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).

3.4 Examples from the Literature

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. 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,*

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

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}*.*

4 Illustration: Binary Choice with Exogenous Characteristics

Typically one will want to allow for endogeneity of at least one component of x_{jt}.However, to illustrate key elements of our approach, we begin with the simple case of binary choice with exogenous x_{jt}. Here we can drop the subscript j, with consumer i selecting the inside good 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 merely states that we consider here the special case of exogenous observables. This assumption is dropped 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)).[17] We relax this assumption in section 6, 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. We then rewrite (5) as

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

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

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

Assumption 3 guarantees that the distribution of $\mu_{it}|t$ is identified from variation in $z_{it}^{(1)}$ within market t. Denote this cumulative 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 know the distribution of $\mu_{it}|t$ for all t, we can recover the value of each ξ_t as well. To see this, let

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

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

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

[17] 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.

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 (9), strict monotonicity of D in ξ_t, and the normalization of ξ_t,

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

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.

Thus far we have shown identification of $F_{\mu_{it}|t}$ and of each latent ξ_t. So for any (x_t, ξ_t) in their support, the value of

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

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

Theorem 1. *Consider the binary choice setting with preferences given by (5). Under Assumptions 1–3, the distribution of u_{it} conditional on any $(x_t, \xi_t, z_{it}) \in \chi$.*

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)).[18] Second, we use variation in choice characteristics across markets 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

[18]See also Matzkin (2007a, 2007b).

(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.

5 Multinomial Choice: Full Identification

Here we consider the general case of multinomial choice with endogenous characteristics using the specification of preferences in (5). 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, essentially requiring sufficient variation in $\left(z_{i1t}^{(1)}, \ldots, z_{iJ_t t}^{(1)}\right)$ to move choice probabilities through the entire unit simplex.[19] 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. However, we will also explore results that do not require this assumption in section 6.

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 exogenous conditioning variables. We will consider two alternative sets of instrumental variables conditions below.

5.1 Identification with Fully Independent Instruments

We first explore identification using instrumental variables conditions from Chernozhukov and Hansen (2005). Here we assume $x_{jt}^{(1)}$ is continuously distributed across j and t, with

[19] This is only "essentially" required by the large support condition because it does not require continuity of choice probabilities in $z_{it}^{(1)}$.

conditional density function $f_{x_j}\left(x_{jt}^{(1)}|\mathrm{w}_{jt}\right)$.[20] For the remainder of this section we will condition on a value of $\left(x_{1t}^{(2)},\ldots,x_{Jt}^{(2)}\right)$, suppress these arguments in the notation, and let x_{jt} now denote $x_{jt}^{(1)}$. This requires that we rewrite (5)

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

i.e., with a j subscript on the functions μ_j, which are different unless $x_{jt}^{(2)}$ is the same for all j.[21] Let

$$\delta_{jt} = D_j\left(x_{jt},\xi_{jt}\right) \equiv med\left[\mu_j\left(x_{jt},\xi_{jt},z_{ijt}^{(2)},\omega_{it}\right)\Big| x_{jt},\xi_{jt}\right] \tag{10}$$

and let $f_{\delta_j}\left(\cdot|x_{jt},\mathrm{w}_{jt}\right)$ denote its density conditional on w_{jt}.[22] Fix some small positive constants $\epsilon_q,\epsilon_f>0$. For each j, for $\tau\in(0,1)$ define $\mathcal{L}_j(\tau)$ to be the convex hull of functions $m_j(\cdot,\tau)$ that satisfy

(a) for all w_{jt}, $\Pr\left(\delta_{jt}\le m_j\left(x_{jt},\tau\right)|\tau,\mathrm{w}_{jt}\right)\in[\tau-\epsilon_q,\tau+\epsilon_q]$; and

(b) for all x in the support of x_{jt}, $m_j(x,\tau)\in s_j(x)\equiv\left\{\delta : f_{\delta_j}(\delta|x,\mathrm{w})\ge\epsilon_f \ \forall \mathrm{w} \text{ with } f_{x_j}(x|\mathrm{w})>0\right\}$.

We assume the following

Assumption 5. $\xi_{jt} \perp\!\!\!\perp \left(\mathrm{w}_{jt},z_{ijt}\right)\forall j,t.$

Assumption 6. For all j, (i) the random variables x_{jt} and δ_{jt} have bounded support;
(ii) for any $\tau\in(0,1)$, for any bounded function $B_j(x,\tau)=m_j(x,\tau)-D_j(x,\tau)$ with $m_j(\cdot,\tau)\in\mathcal{L}_j(\tau)$ and $\varepsilon_{jt}\equiv\delta_{jt}-D_j\left(x_{jt},\tau\right)$, $E\left[B_j\left(x_{jt},\tau\right)\psi_j\left(x_{jt},\mathrm{w}_{jt},\tau\right)|\mathrm{w}_{jt}\right]=0$ a.s. only if $B_j\left(x_{jt},\tau\right)=0$ a.s., where $\psi_j\left(x,\mathrm{w},\tau\right)=\int_0^1 f_{\varepsilon_j}\left(\sigma B_j(x,\tau)|x,\mathrm{w}\right)\,d\sigma$.
(iii) the density $\int_0^1 f_{\varepsilon_j}(e|x,\mathrm{w})$ of ϵ_{jt} is bounded and continuous in e on $\mathbb{R}$;
(iv) $D_j(x,\xi)\in s_j(x)$ for all (x,ξ) in their support.

[20] This could be dropped by appealing below to Theorems 2 and 3 (and the associated rank conditions) in Chernozhukov and Hansen (2005) instead of their Theorem 4. We consider the case of a continous endogenous characteristic here because price is our leading example.

[21] Recall that $x_{jt}^{(2)}$ may include product dummies, so in general the functions μ_j and $\mu_{j'}$ need not have any particular relation.

[22] 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.

Assumption 5 is a strong exclusion restriction. Assumption 6 is a particular type of "bounded completeness" condition (Chernozhukov and Hansen, 2005, Appendix C), ensuring that the instruments induce sufficient variation in the endogenous variables. This condition plays the role of the standard rank condition for linear models, but for the nonparametric nonseparable model $\delta_j = D_j(x,\xi)$.[23] With these assumptions, we obtain the following result.

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 $\left(z_{i1t}^{(2)}, \ldots, z_{iJt}^{(2)}\right)$. Let $\mu_{ijt} = \mu_j\left(x_{jt}, \xi_{jt}, z_{ijt}^{(2)}, \omega_{it}\right)$ and observe that

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

Holding t fixed, $\mu_{ijt} \perp\!\!\!\perp z_{ijt}^{(1)}$ by Assumption 1 and our conditioning on $z_{it}^{(2)}$. Assumption 4 then guarantees identification of the marginal distribution of each $\mu_{ijt}|t$ for each j. This implies, for each j, identification of the conditional median

$$\delta_{jt} = med\left[\mu\left(x_{jt}, \xi_{jt}, \omega_{it}\right)|x_{jt}, \xi_{jt}\right] = med\left[\mu\left(x_{jt}, \xi_{jt}, \omega_{it}\right)|t\right] \tag{11}$$

Thus, the left side of (10) can be treated as observed. Further, the function D_j in (10) must be strictly increasing in ξ_{jt}. Under Assumptions 5 and 6, Theorem 4 of Chernozhukov and Hansen (2005) then implies that each function D_j (and therefore each realization ξ_{jt}) is identified. Finally, observe that for any market t

$$\begin{aligned} p_{i0t} &= \Pr\left(z_{i1t}^{(1)} + \mu_{i1t} < 0, \ldots, z_{iJt}^{(1)} + \mu_{iJt} < 0 \middle| t, z_{i1t}^{(1)}, \ldots, z_{iJt}^{(1)}\right) \\ &= \Pr\left(\mu_{i1t} < -z_{i1t}^{(1)}, \ldots, \mu_{iJt} < -z_{iJt}^{(1)} \middle| t, z_{i1t}^{(1)}, \ldots, z_{iJt}^{(1)}\right) \end{aligned} \tag{12}$$

[23] Chernozhukov and Hansen (2005) discuss sufficient conditions. We also consider an alternative to Assumption 6 below.

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 $\left\{(x_{jt}, z_{ijt}, \xi_{jt})\right\}_{j\in\mathcal{J}_t}$ in their support given $\mathcal{J}_t$. Since $u_{ijt} = z_{ijt}^{(1)} + \mu_{ijt}$, the result follows. □

5.2 Identification with Mean-Independent Instruments

A possible limitation of Theorem 2 is that Assumption 6 may be difficult to check and/or 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.

Conditioning on $x_t^{(2)}$ as in the prior section, suppose (for this subsection only) that each consumer i's conditional indirect utilities can be represented by

$$\tilde{u}_{ijt} = \beta_{it} z_{ijt}^{(1)} + \tilde{\mu}_j\left(x_{jt}, z_{ijt}^{(2)}, \omega_{it}\right) + \gamma_{it}\xi_{jt} \qquad j = 1, \ldots, \mathcal{J}_t \tag{13}$$

where β_{it} is strictly positive with probability one and the expectations $E\left[\beta_{it}\right]$, $E\left[\gamma_{it}\right]$, and $E\left[\tilde{\mu}_j\left(x_{jt}, z_{ijt}^{(2)}, \omega_{it}\right)|x_{jt}, z_{ijt}^{(2)}\right]$ are finite. This imposes a restriction relative to (4) but is still quite general relative to the prior literature. A representation of preferences equivalent to (13) is

$$u_{ijt} = z_{ijt}^{(1)} + \mu_j\left(x_{jt}, \xi_{jt}, z_{ijt}^{(2)}, \omega_{it}\right) \qquad \forall i, j = 1, \ldots, \mathcal{J}_t \tag{14}$$

where now

$$\mu_j\left(x_{jt}, \xi_{jt}, z_{ijt}^{(2)}, \omega_{it}\right) = \frac{\tilde{\mu}_j\left(x_{jt}, z_{ijt}^{(2)}, \omega_{it}\right)}{\beta_{it}} + \frac{\gamma_{it}}{\beta_{it}}\xi_{jt}. \tag{15}$$

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

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

and scale normalization

$$E\left[\frac{\gamma_{it}}{\beta_{it}}\right] = 1. \tag{16}$$

Both are without further loss of generality. The latter, for example, defines units of the unobservables ξ_{jt}.

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

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

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 (11). Recall that in that proof we fixed the value of $\left(z_{i1t}^{(2)}, \ldots, z_{iJt}^{(2)}\right)$. With the separable structure (15) and the normalization (16), for each j we now we let

$$\delta_{jt} = E\left[\mu_j\left(x_{jt}, \xi_{jt} z_{ijt}^{(2)}, \omega_{it}\right)\middle| t\right] = D_j\left(x_{jt}\right) + \xi_{jt}$$

for an unknown function D_j As in the proof of Theorem 2, 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 each function D_j from observation of $(\delta_{jt}, x_{jt}, \tilde{\mathrm{w}}_{jt})$ (Newey and Powell (2003)).

Assumption 8. For all j and all functions $B_j\left(x_{jt}\right)$ with finite expectation, $E\left[B_j\left(x_{jt}\right)|\mathrm{w}_{jt}\right] = 0$ a.s. implies $B_j\left(x_{jt}\right) = 0$ a.s.

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

Theorem 3. *Under the representation of preferences (14), suppose Assumptions 1, 4, and 7 hold. Then the joint distribution of* $\{u_{ijt}\}_{j\in\mathcal{J}_t}$ *conditional on any* $\left(\mathcal{J}_t, \left\{(x_{jt}, z_{ijt}, \xi_{jt})\right\}_{j\in\mathcal{J}_t}\right)$

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 each D_j and therefore of each ξ_{jt}. The remainder of the proof then follows that of Theorem 2 exactly, beginning with (12). □

The completeness condition here (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 $D_j(x_{jt})$ at all points x_{jt}.[24]

6 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 $\left(z_{i1}^{(1)}, \ldots, z_{iJ_t}^{(1)}\right)$ 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 exogenous observables 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, one would like to understand how heavily the results rely on the tails of the large support and what can be learned from more limited variation. We explore these questions here.

We show that more limited variation is 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. We also show continuity of the identified features with respect to the support of the micro data. In particular, moving from our limited support condition to the full support condition moves

[24] 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 $f_j(x_{jt})$ such that $E[\ f_j(x_{jt})] = 0$ and $var(f_j(x_{jt})) > 0$, there exists a function $h_j(\cdot)$ such that $f_j(x_{jt})$ and $h_j(\mathrm{w}_{jt})$ 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 $\left(\hat{x}^1, \ldots, \hat{x}^K\right) \times \left(\hat{\mathrm{w}}^1, \ldots, \hat{\mathrm{w}}^L\right)$, completeness corresponds to a full rank condition on the matrix $\{\sigma_{kl}\}$ where $\sigma_{kl} = \Pr(x_{jt} = \hat{x}^k | \mathrm{w}_{jt} = \hat{\mathrm{w}}^l)$.

the identified features of the model smoothly toward the full identification results of the preceding section.

For multinomial choice we obtain these results under a somewhat more restrictive specification of preferences than that in (5). Up to this qualification, however, these results should be a comforting. Demand is identified without the large support condition. And although we require the large support for full identifiability of the random utility model in the previous section, the identification is not knife-edge: the tails of the large support are needed only to determine the tails of the joint distributions of utilities.

6.1 Binary Choice

6.1.1 Identification of Demand

As before, we begin with binary choice to illustrate our main insights. We begin with the relaxed support condition on $z_{it}^{(1)}$, requiring a single "common choice probability" that is attainable in all markets by the appropriate choice of $z_{it}^{(1)}$ in its support.

Assumption 9. For some $\tau \in (0,1)$, for every market t there exists a unique $z_t^\tau \in \text{supp } z_{it}^{(1)}$ 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.[25] This is not innocuous but is much less demanding than the full support condition.

In the case of binary choice we obtain results using the general specification of preferences in (5).[26] Here, the consumer chooses the inside good whenever (fixing $z_{it}^{(2)}$ and suppressing it)

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

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

[26] In the case of binary choice, the additive separability in $z_{it}^{(1)}$ is without loss if the utility for the inside good is strictly increasing in $z_{it}^{(1)}$. See Berry and Haile (2008a) for additional results for the special case of binary choice and threshold crossing models.

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

$$z_t^{\tau}=\zeta\left(x_t,\xi_t;\tau\right)\tag{17}$$

for some function $\zeta\left(\cdot;\tau\right)$ 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 $\zeta\left(\cdot;\tau\right)$, and therefore of each ξ_t, follows from (17) as in the preceding sections, using the nonparametric instrumental variables result of Chernozhukov and Hansen (2005). This requires the same type of bounded completeness assumption made in section 5.1. We state this formally as Assumption 13 in the Appendix. With each ξ_t known, the observable choice probabilities reveal the structural choice probabilities

$$\rho\left(x_t,\xi_t,z_{it}\right)=\Pr\left(y_{it}=1|x_t,\xi_t,z_{it}\right)\tag{18}$$

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 (5), suppose Assumptions 1, 5, 9, and 13 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.*

6.1.2 Continuity of the Identified Features

Theorem 4 required 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 function $\zeta\left(\cdot;\tau\right)$ for each common choice probability τ, each then determining the τth quantile of $-\mu(x_t, \xi_t, \omega_{it})|x_t, \xi_t$. Since

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

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" described above and shows the tails of $z_{ijt}^{(1)}$ under the large support assumption are used only to identify the tails of the conditional distributions of utilities.

6.2 Multinomial Choice

For multinomial choice we will require a different representation of preferences:

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

where μ is assumed strictly increasing in its first argument. This is similar to (13) in requiring that $z_{ijt}^{(1)}$ and ξ_{jt} be perfectly substitutable. Here we require all consumers to have the same marginal rate of substitution between $z_{ijt}^{(1)}$ and ξ_{jt}, but we allow the index $z_{ijt}^{(1)} + \xi_{jt}$ to enter the utility function in a fully nonparametric way.

A key implication of (19) is that choice probabilities depend on the sums

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

rather than on each $z_{ijt}^{(1)}$ and ξ_{jt} separately. Letting $\boldsymbol{\lambda}_t = (\lambda_{i1t}, \ldots, \lambda_{iJt})$, $\mathbf{x}_t = (x_{1t}, \ldots, x_{Jt})$, and $\mathbf{z}_{it}^{(2)} = \left(z_{i1t}^{(2)}, \ldots, z_{iJt}^{(2)}\right)$, we can then write the structural choice probabilities as

$$\rho_j\left(\boldsymbol{\lambda}_t, \mathbf{x}_t, \mathbf{z}_{it}^{(2)}\right).$$

Following Gandhi (2008), we make the following "strong substitutes" assumption.

Assumption 10. Consider any $\left(\mathbf{x}_t, \mathbf{z}_{it}^{(2)}\right)$ and any $\boldsymbol{\lambda}_t$ such that $\rho_j\left(\boldsymbol{\lambda}_t, \mathbf{x}_t, \mathbf{z}_{it}^{(2)}\right) > 0$ for all $j \in \mathcal{J}$. For any strict subset $\mathcal{K} \subset \mathcal{J}$, there exists $k \in \mathcal{K}$ and $j \notin \mathcal{K}$ such that $\rho_j\left(\boldsymbol{\lambda}_t, \mathbf{x}_t, \mathbf{z}_{it}^{(2)}\right)$ is strictly decreasing in λ_{kt}.

Given the monotonicity of u_{ijt} in λ_{jt}, this is a natural regularity condition requiring that in every binary partition of $\mathcal{J}$, there is some substitution between cells of the partition. This is guaranteed if there are always consumers on the margin of indifference between every pair of choices. This holds in standard models like the random coefficients multinomial logit but is stronger than necessary. For example, in a pure vertical model a given product substitutes with at most two others, yet the condition holds.[27]

Let $\triangle^J$ denote the $J-1$ dimensional unit simplex. With Assumption 10 we can follow the argument used to prove Theorem 2 of Gandhi (2008) to show the following lemma, which generalizes the well-known invertibility results for linear discrete choice models from Berry (1994).[28]

Lemma 1. *Consider any choice probability vector* $\mathbf{p} = (p_1, \ldots, p_J)'$ *on the interior of* $\triangle^J$. *Under Assumptions 1 and 10, for any* $\left(\mathbf{x}_t, \mathbf{z}_{it}^{(2)}\right)$ *there is at most one vector* $\boldsymbol{\lambda} \in \mathbb{R}^J$ *such that* $\rho_j\left(\boldsymbol{\lambda}, \mathbf{x}_t, \mathbf{z}_{it}^{(2)}\right) = p_j$ *for all* j .

Proof. Fix $\left(\mathbf{x}_t, \mathbf{z}_{it}^{(2)}\right)$ and suppress them in the notation. Now suppose, contrary to the claim, that for some $\boldsymbol{\lambda} \neq \boldsymbol{\lambda}'$, $\rho_j(\boldsymbol{\lambda}) = \rho_j(\lambda') = p_j$ for all j. Note that since we have normalized the utility of the outside good to zero for all choice sets, we can define $\boldsymbol{\lambda}_0 = \boldsymbol{\lambda}_0' = 0$ as a

[27] For example, consider a 5 good vertical model and $\mathcal{K} = \{2, 3, 4\}$. Good 3 does not substitute with goods outside of $\mathcal{K}$, but goods 2 and 4 do. Thus the condition holds. It requires only that there be some element of $\mathcal{K}$ that substitutes with some good outside $\mathcal{K}$. This will be true here for any proper subset $\mathcal{K} \subset \mathcal{J} = \{0, 1, \ldots, 5\}$.

[28] See Berry and Pakes (2007) for an alternative set of sufficient conditions. Berry (1994) and Berry and Pakes (2007) show existence and uniqueness of an inverse choice probability in models with an additively separable δ_{jt}. Gandhi (2008) relaxes the separability requirement. Our lemma addresses only uniqueness conditional on existence. Under our maintained assumption that the model is correctly specified, given any observed choice probability vector, there must exist a vector $(\delta_1, \ldots, \delta_J)$ that rationalizes it. Gandhi (2008) provides conditions gauranteeing that an inverse exists for every choice probability vector in $\triangle^J$. Our uniqueness result differs from his only slightly, mainly in recognizing that the argument applies to a more general model of preferences.

notational convention without loss. Without loss, let $\lambda'_j > \lambda_j$ for some $j \in \mathcal{J}$. Because $0 \in \mathcal{J}$, there must then exist a strict subset of choices $\mathcal{K} \subset \mathcal{J}$ such that $\lambda'_j > \lambda_j \forall j \in \mathcal{K}$ and $\lambda'_j \leq \lambda_j \forall j \in \mathcal{J} - \mathcal{K}$. For this subset $\mathcal{K}$ let $k \in \mathcal{K}$ be the index of a product referred to (as "k") in Assumption 10. Now define a new vector $\boldsymbol{\lambda}^*$ by

$$\begin{aligned} \lambda^*_k &= \lambda'_k \\ \lambda^*_j &= \lambda_j \forall j \neq k. \end{aligned}$$

Monotonicity of u_{ijt} in δ_{jt} implies that $\rho_j(\boldsymbol{\lambda}^*) \leq \rho_j(\boldsymbol{\lambda})$ for all $j \in \mathcal{J} - \mathcal{K}$. Furthermore, Assumption 10 implies

$$\sum_{j \in \mathcal{J} - \mathcal{K}} \rho_j(\boldsymbol{\lambda}^*) < \sum_{j \in \mathcal{J} - \mathcal{K}} \rho_j(\boldsymbol{\lambda})$$

so that (since probabilities must sum to one)

$$\sum_{j \in \mathcal{K}} \rho_j(\boldsymbol{\lambda}^*) > \sum_{j \in \mathcal{K}} \rho_j(\boldsymbol{\lambda}).$$

But then by monotonicity of u_{ijt} in λ_{jt}, we have

$$\sum_{j \in \mathcal{K}} \rho_j(\boldsymbol{\lambda}') \geq \sum_{j \in \mathcal{K}} \rho_j(\boldsymbol{\lambda}^*) > \sum_{j \in \mathcal{K}} \rho_j(\boldsymbol{\lambda})$$

which contradicts the hypothesis $\rho_j(\boldsymbol{\lambda}) = \rho_j(\boldsymbol{\lambda}') = p_j$ for all j. $\square$

Finally, we generalize the previous common choice probability assumption in the natural way.

Assumption 11. For all $\mathcal{J}_t$, there exists $q = (q_0, q_1, \ldots, q_J) \in \triangle^{J_t}$ such that for every market t there is a unique vector $\mathbf{z}^q_t = (z^q_{1t}, \ldots, z^q_{1t}) \in \text{supp}\left(z^{(1)}_{i1t}, \ldots, z^{(1)}_{iJ_t t}\right)$ such that $q_j = \Pr(y_{it} = j \mid x_{1t}, \ldots, x_{J_t t}, z_{i1t}, \ldots, z_{iJ_t t})_{\mathbf{z}^{(1)}_{it} = \mathbf{z}^q_t}$ for all $j = 1, \ldots, J_t$.

If $\mu\left(z^{(1)}_{ijt} + \xi_{jt}, x_{jt}, z^{(2)}_{ijt}, \omega_{it}\right)$ is continuously distributed conditional on $\left(z^{(1)}_{ijt} + \xi_{jt}, x_{jt}, z^{(2)}_{ijt}\right)$, uniqueness of $\mathbf{z}^q_t$ is guaranteed by Lemma 1. Beyond this, the requirement of Assumption 11 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. Note that the value of q satisfying this condition need not be known *a priori*, since this is observable. Indeed, the existence of the common choice probability is directly testable. This condition, while still demanding sufficient J-dimensional micro data, is clearly weaker than the full support condition, which essentially requires all elements of $\triangle^{J_t}$ to be common choice probabilities.

Theorem 5. *In the multinomial choice model with preferences given by (19), suppose Assumptions 1, 7 , 8, 10, and 11 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. Fixing the vector $\mathbf{z}_t^{(2)} = \hat{\mathbf{z}}_t^{(2)}$, suppressing it in the notation, and letting $\mu_j\left(\lambda_{ijt}, x_{jt}, \omega_{it}\right) = \mu\left(\lambda_{ijt}, x_{jt}, \hat{z}_{ijt}^{(2)}, \omega_{it}\right)$, we have

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

Fix $\mathbf{x}_t = (x_{1t}, \ldots, x_{J_t t})$ and let q be the common choice probability vector. From Lemma 1 and Assumption 11, there is a unique vector $\boldsymbol{\lambda}\left(\mathbf{x}_t, q\right) = \left(\lambda_1\left(\mathbf{x}_t, q\right).,\ldots, \lambda_{J_t}\left(\mathbf{x}_t, q\right)\right)$ such that

$$\rho_j\left(\boldsymbol{\lambda}\left(\mathbf{x}_t, q\right), \mathbf{x}_t\right) = q_j \quad \forall j.$$

Further, by the definition of $\mathbf{z}_t^q$ and $\lambda_j\left(\mathbf{x}_t, \mathbf{q}\right)$, $\lambda_j\left(\mathbf{x}_t, \mathbf{q}\right) = \xi_{jt} + z_{jt}^q$, so that

$$z_{jt}^q = \lambda_j\left(\mathbf{x}_t, q\right) - \xi_{jt} \quad \forall j, t. \tag{20}$$

The equations (20) 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 identification result in Newey and Powell (2003) for nonparametric regression with instrumental variables. As demonstrated above, knowledge of all ξ_{jt} identifies the structural choice probability functions (demand). □

7 Testable Restrictions

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

The assumption of a vertical $z_{ijt}^{(1)}$ has immediate implications for the observed conditional choice probabilities.

Remark 1. *(i) Suppose preferences can be characterized by (5) with $z_{ijt}^{(1)}$ independent of ξ_{jt}. Then under Assumption 1,* $\Pr\left(y_{it}=j|\mathcal{J}_t,\{x_{jt},w_{jt},z_{ikt}\}_{k\in\mathcal{J}_t}\right)$ *is increasing in $z_{ijt}^{(1)}$. (ii) Suppose preferences can be characterized by (5) or by (19) with μ strictly increasing in its first argument. Then* $\Pr\left(y_{it}=j|t,\{z_{ikt}\}_{k\in\mathcal{J}_t}\right)$ *is increasing in $z_{ijt}^{(1)}$.*

The first restriction involves variation in choice probabilities across markets and depends on exogeneity of ξ_{jt}. The second addresses variation within a market, where $\left\{\xi_{jt}\right\}_{j\in\mathcal{J}_t}$ are held fixed. Both implications are immediate from the requirement that the utility from good j be strictly increasing in $z_{ijt}^{(1)}$. Furthermore, it is clear that the restriction need not hold if utilities sometimes are decreasing in $z_{ijt}^{(1)}$.

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 market t. As usual, we condition on $z_{it}^{(2)}$ and suppress it in the notation.

Remark 2. *In the binary choice model with preferences given by (5), suppose Assumptions 1, 5, 13, and 9 hold. Then $\xi_t\left(z_t^\tau;\tau,x_t\right)$ must be strictly decreasing in z_t^τ across markets.*

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

$$\Phi\left(\frac{z_t^\tau}{\xi_t}\right) = \tau \quad \forall t. \tag{21}$$

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 2.*

The restriction in Remark 2 follows from the requirement of a vertical ξ_{jt}. An additional restriction is implied by the restriction to a scalar choice/market-specific unobservable.

Remark 3. *In the binary choice model with preferences given by (5), suppose Assumptions 1, 5, and 13 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,* $\sigma \neq \sigma'$.

8 Extension: Aggregate Data with Market Groups

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). In Berry and Haile (2008b) we explore identification in such settings. 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

$$u_{ijt} = x_{jt}^{(i)} + \mu(x_{jt}^{(ii)}, \xi_{jt}, \omega_{it}). \tag{22}$$

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 12. 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 12 is different from the parallel large support Assumption 4. Here we require sufficient variation in a special product characteristic rather than an special consumer-product characteristic. The role of this assumption is the same, however: to trace out the distribution of the random component of (22) within each market group.

Now the setup is isomorphic to that in section 5. 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 5. 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.[29]

[29] An interesting question is what can be learned in a single market with a large choice set, i.e., with $J \to \infty$ (see Berry, Linton, and Pakes (2004)). Suppose that x_{jt} does not include product dummies but preferences can still be represented by 5, imposing a symmetry condition that the same function μ apply to all products. Fixing a market with a finite choice set, the market share of the outside good is

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

A large support condition would give identification of the joint distribution of $\left(\mu\left(x_1, \xi_1, \omega_{it}\right), \ldots, \mu\left(x_J, \xi_J, \omega_{it}\right)\right)$, so that each $\delta_j \equiv med\ \mu\left(x_j, \xi_j, \omega_{it}\right) | x_j, \xi_j$ could be considered known. Since we can write $\delta_j = D\left(x_j, \xi_j\right)$, letting $J \to \infty$, it may be possible to extend the identification result of Chernozhukov and Hansen (2005) to obtain identification of D, which would then imply identification of each ξ_j, with full identification then following as above.

9 Conclusion

We have studied nonparametric identification of models of multinomial choice demand, allowing for choice-specific unobservables, endogenous choice characteristics, and arbitrary random heterogeneity across consumers in tastes for products and/or characteristics. We obtained full identification using the same kind of large support assumption used to show identification in semiparametric models, and the same instrumental variables conditions required for identification of nonparametric regression models. Further, the results rely on the large support only for identification of tail probabilities, and identification of demand holds under a significantly weaker support condition.

While one goal of our work has been to obtain results with few restrictions on preferences, there are some costs to a choice not to place more structure on the form of utility functions. One is that some types of counterfactuals will not be identifiable.[30] 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, but is inherent to a nonparametric analysis: extrapolation and interpolation typically require some parametric structure. Of course, one may have more confidence in extrapolations when identification holds nonparametrically within the support of the data generating process.

A second limitation concerns welfare. Our model (5) incorporates quasilinear preferences and can therefore be used to characterize changes in utilitarian social welfare (in aggregate, or across subpopulations defined by observables).[31] However, it 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

[30] The model enables identification of some counterfactuals outside the support of the data generating process—for example, removal of a product from the choice set.

[31] 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.

after. This is because our model specifies a distribution of conditional indirect utilities, not a distribution of parameters whose realizations can be associated with an individual. This points out a limitation of nonparametric random utility models as a theoretical foundation for some kinds of welfare analysis: such welfare calculations require additional *a priori* structure.

An example of a model with sufficient structure to address these welfare questions (and to extrapolate/interpolate) is the linear random coefficients random utility model (Example 1)

$$u_{ijt} = x_{jt}\beta_{it} + z_{ijt}\gamma + \xi_{jt} + \epsilon_{ijt}. \tag{23}$$

This generates a special case of our model, so we have 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. However, different joint distributions of $\left(\beta_{it}, \{\epsilon_{ijt}\}_{j\in\mathcal{J}}\right)$ that imply the same conditional joint distributions of utilities need not have the same implications for welfare or extrapolation/interpolation. Going from our results to identification of the distribution of parameters in (23) is 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 beyond those required for our results. Whether this enables any relaxation of existing identification results for linear random coefficients models (e.g., Ichimura and Thompson (1998), Briesch, Chintagunta, and Matzkin (2005), Gautier and Kitamura (2007), Fox and Gandhi (2008)) is an open question.

Finally, while a novel aspect of our work is it examination of identification without large support conditions, even our weaker "common choice probability" condition requires J-dimensional micro data. One can easily imagine applications where this will not be available. When no micro data are available, one is in the case of market-level data. We explore that case in Berry and Haile (2008b). Whether the conditions for identification we identify there could be relaxed in intermediate cases—where there is some micro data, but of a lower dimension than that of the choice set—is an interesting question for future work.

Appendix

Here we state Assumption 13, used in Theorem 4. From equation (17) we have

$$z_t^{\tau} = \zeta(x_t, \xi_t; \tau)$$

where x_t denotes the endogenous characteristic of choice 1. For given τ, let $f_{z^\tau}(\cdot|x_t, \mathrm{w}_t)$ denote the density of conditional z_t^τ on x_t and the instruments w_t. 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

(a) for all w_t, $\Pr(z_t^\tau \leq m(x_t, \tau)|\tau, \mathrm{w}_t) \in [\tau - \epsilon_q, \tau + \epsilon_q]$; and

(b) for all x in the support of x_t, $m(x, \tau) \in s(x) \equiv \{\delta : f_\delta(\delta|x, \mathrm{w}) \geq \epsilon_f \ \forall \mathrm{w} \text{ with } f_x(x|\mathrm{w}) > 0\}$.

Assumption 13. (i) The random variables x_t and z_t^τ have bounded support.

(ii) For any $\tau \in (0,1)$, for any bounded function $B(x,\tau) = m(x,\tau) - \zeta(x,\tau)$ with $m(\cdot,\tau) \in \mathcal{L}(\tau)$ and $\varepsilon_t \equiv z_t^\tau - \zeta(x_t, \tau)$, $E[B(x_t,\tau)\psi(x_t, \mathrm{w}_t, \tau)|\mathrm{w}_t] = 0$ a.s. only if $B(x_t, \tau) = 0$ a.s., where $\psi(x, \mathrm{w}, \tau) = \int_0^1 f_\varepsilon(\sigma B(x,\tau)|x, \mathrm{w}) \ d\sigma$.

(iii) the density $\int_0^1 f_\varepsilon(e|x, \mathrm{w})$ of ϵ_t is bounded and continuous in e on $\mathbb{R}$;

(iv) $\zeta(x, \xi) \in s(x)$ for all (x, ξ) in their support.

References

Altonji, J., and R. L. Matzkin (2005): "Cross-Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors," *Econometrica*, 73, 1053–1102.

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.

Barbera, S., and P. K. Pattanaik (1986): “Falmagne and the Rationalizability of Stochastic Choices in Terms of Random Orderings,” *Econometrica*, 54, 707–715.

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 (2008a): “Identification of a Nonparametrc Threshold Crossing Model,” Discussion paper, Yale University.

——— (2008b): “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.

Chintagunta, P. K., D. C. Jain, and N. J. Vilcassim (1991): "Investigating Heterogeneity In Brand Preferences In Logit Models for Panel Data," *Journal of Marketing Research*, 28, 417–428.

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.

Hoderlein, S. (2008): "Endogeneity in Semiparametric Binary Random Coefficient Models," Discussion paper, Brown 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.

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.

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

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

——— (1987): "Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data," *Econometrica*, 55, 357–362.

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

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

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

——— (2004): "Unobservable Instruments," Discussion paper, Northwestern University.

——— (2005): "Identification in Nonparametric Simultaneous Equations," Discussion paper, Northwestern University.

——— (2007a): "Heterogeneous Choice," in *Advances in Economics and Econometrics, Theory and Applications, Ninth World Congress of the Econometric Society*, ed. by R. Blundell, W. Newey, and T. Persson. Cambridge University Press.

——— (2007b): "Nonparametric Identification," in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. Leamer, vol. 6B. Elsevier.

——— (2008): "Identification in Nonparametric Simultaneous Equations," *Econometrica*, 76, 945–978.

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.

Petrin, A., and K. Train (2009): "A Control Function Approach to Endogeneity in Consumer Choice Models," *Journal of Marketing Research*, forthcoming.

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*.