NBER WORKING PAPER SERIES

NONPARAMETRIC IDENTIFICATION OF MULTINOMIAL CHOICE DEMAND MODELS WITH HETEROGENEOUS CONSUMERS

Steven T. Berry
Philip A. Haile

Working Paper 15276
http://www.nber.org/papers/w15276

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2009

We have had helpful conversations on this topic with Liran Einav, Jin Hahn, Hide Ichimura, Jon Levin, Rosa Matzkin and Yuichi Kitamura. We also received useful comments from Sunyoung Park and participants in several conferences and seminars. Financial support from the NSF is gratefully acknowledged. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers

Steven T. Berry and Philip A. Haile

NBER Working Paper No. 15276
August 2009
JEL No. C35

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.

Steven T. Berry
Yale University Department of Economics
Box 208264
37 Hillhouse Avenue
New Haven, CT 06520-8264
and NBER
steven.berry@yale.edu

Philip A. Haile
Department of Economics
Yale University
37 Hillhouse Avenue
P.O. Box 208264
New Haven, CT 06520
and NBER
philip.haile@yale.edu

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. 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 goods in a wide range of applied fields of economics and related disciplines. Examples include transportation and urban economics (e.g., Domencich and McFadden (1975)), industrial organization (e.g., Berry, Levinsohn, and Pakes (1995)), international trade (e.g., Goldberg (1995)), marketing (e.g., Guadagni and Little (1983)), education (e.g., Manski and Wise (1983)), migration (e.g., Schultz (1982)), voting (e.g., Poole and Rosenthal (1985)), and health economics (e.g., Capps, Dranove, and Satterthwaite (2003)). We focus in particular on models in the spirit of Berry (1994), Berry, Levinsohn and Pakes (1995, 2004), Nevo (2001), Petrin (2002) and a large related literature. These models combine two essential features: heterogeneous tastes for choice characteristics, and endogeneity through market/choice-specific unobservables. Although this class of models is used in a wide range of applications, their identification has not been addressed in the prior literature. Without an understanding of the sources of identification in these models, 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 market/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 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. Our formulation not only enables us to consider a very general model of preference heterogeneity, but also leads us to focus directly on identification of the conditional joint distribution of utilities. The advantage of this approach might be unexpected: a natural intuition is that added structure on the way randomness enters the model 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), even the standard linear random coefficients model will involve taste shocks of larger dimension unless significant restrictions are placed on the correlation structure. Focusing directly on the joint distribution of utilities naturally leads to primitives whose dimension exactly matches the dimension of the observables without imposing strong distributional or functional form restrictions.

A second key aspect of our work is our explicit modeling of market/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 market/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 characterization of 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.[1]

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 two standard ideas. One is the use of variation in exogenous 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. We show that this strategy is particularly useful in a micro data setting, where one can exploit variation in individual-level observables within a market, holding market-level unobservables fixed. As usual, we require the observables tracing out the joint distribution to have dimension as large as the choice set. However, we show that the support conditions commonly used with this strategy can be substantially relaxed without losing identification of demand. A second standard idea is the use of exogenous variation in choice sets to decompose variation in the distribution of utilities into the contributions

[1] Although this is the type of endogeneity typically relevant to demand applications, it obviously is not the only possibility. For example, our model typically would not be appropriate for evaluating treatment effects on multinomial outcomes when treatments depend on individual-specific unobservables. See Berry and Haile (2009b) for additional discussion and examples in the context of a generalized regression model.

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). To our knowledge, the applicability of these results to discrete choice settings has not been previously exploited. An implication of our results is that the primary requirement for identification of demand is the availability of instruments. As discussed below, one advantage of micro data is the variety of instruments that may be available.

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 illustrate the robustness of the full identification results to the relaxation of support conditions. Section 7 discusses testable restrictions of key maintained hypotheses. In section 8 we show how our results can be reinterpreted to show identification in one type of environment in which only market level data are available. We conclude in section 9.

2 Relation to the Literature

Our work is motivated by a large applied literature using models of discrete choice demand incorporating two key components: heterogeneous preferences for choice characteristics and endogeneity through market/choice-specific unobservables. The former allows a flexible model of substitution patterns (e.g., cross-elasticities),[2] while the latter is essential if one

[2]See, e.g., the discussions in Domencich and McFadden (1975), Hausman and Wise (1978) and Berry, Levinsohn, and Pakes (1995). Early models of discrete choice with heterogeneous tastes for characteristics

is to control for the classical endogeneity of prices. Although there is a large literature on identification of discrete choice models,[3] there has been no analysis that covers even the linear version of these models typically used in the applied literature. Thus, although we relax functional form restrictions considered in the prior identification literature, a more fundamental distinction is that our model allows simultaneously for heterogeneity in preferences for choice characteristics and endogeneity through market/choice-specific unobservables.

Identification of heterogeneous preferences for choice characteristics has been explored by Ichimura and Thompson (1998) and Briesch, Chintagunta, and Matzkin (2005). Ichimura and Thompson (1998) studied the linear random coefficients binary choice model. Briesch, Chintagunta, and Matzkin (2005) considered multinomial choice, allowing generalizations of the linear random coefficients model. Neither considers endogenous choice characteristics, and both impose restrictions on how heterogeneity enters that we will not require.

Lewbel (2000) considered identification in the semi-parametric linear random utility model

$$v_{ij} = x_{ij}\beta + \epsilon_{ij} \tag{1}$$

where the distribution of ϵ_{ij} can vary with x_{ij}. Unlike (1), our model makes a distinction between the unobservables responsible for taste heterogeneity and those responsible for endogeneity.[4] To see why this is essential in applications to discrete choice demand, suppose we wish to describe how quantities would change in response to an exogenous change in the price of good j—e.g., to describe own- and cross-price demand elasticities. Accounting for heterogeneity in tastes requires that the entire distribution of v_{ij} (not just its mean) be permitted to change with the change in price. At the same time, controlling for the endogeneity of price requires that all else (in particular, market/product-level unobservables that are correlated with price) be held fixed. Meeting both requirements is not possible

include those in Quandt (1966), Quandt (1968), and Domencich and McFadden (1975).

[3] Important early work includes Manski (1985, 1987, 1988) and Matzkin (1992, 1993).

[4] We also relax functional form restrictions; for example, we do not require mean effects of observables to enter separably from unobservables.

in a model like (1) that incorporates a single "composite" error term. Thus, while (1) provides an attractive structure if the parameter β is the object of interest, the model lacks the structure needed to define key objects of interest in the context of discrete choice demand.

Honoré and Lewbel (2002) consider a binary panel version of the model in Lewbel (2000), relying on linearity in a composite error term and focusing on identification of a slope parameter. Altonji and Matzkin (2005) consider a similar but nonparametric model. Their results for discrete choice models focus on identification and estimation only of local average responses. Other work considering models similar to (1) includes Hong and Tamer (2004), Blundell and Powell (2004), Lewbel (2005), and Magnac and Maurin (2007).

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/heterogeneous tastes for characteristics.[5] Hoderlein (2008) allows for both heteroskedasticity and endogeneity in the case of a binary choice model, focusing on identification of a particular average derivative. Hoderlein (2008), like Blundell and Powell (2004) and Matzkin (2004), limit attention to binary choice in semiparametric triangular models, leading to the applicability of control function methods or the related idea of "unobserved instruments."[6] For binary choice demand, triangular models can be appropriate when price depends on either a demand shock or a cost shock, but not both. In the case of multinomial choice, standard oligopoly models imply that each price depends on the entire vector of demand shocks (and cost shocks, if any). Thus, triangular models do not permit the type of endogeneity relevant for most applications to discrete choice demand.[7]

Berry and Haile (2009a) considers identification of multinomial choice models allowing heterogeneity and endogeneity when only market level data are available. There we explore extensions of ideas developed in this paper as well as a change-of-variables argument with

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

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

[7] Gautier and Kitamura (2007) have considered binary choice in a linear random coefficients environment. They include a brief discussion of endogeneity under a triangular structure.

close connections to control function methods and to classical supply and demand analysis. Lack of micro data makes it impossible to exploit the within-market variation that plays a central role throughout the present paper, and results are obtained there under more restrictive representations of preferences than those considered here. Berry and Haile (2009a) includes results for identification of marginal costs under a variety of oligopoly models. Those results extend immediately to the micro data environment considered here, given the results on identification of demand below.

Both Berry and Haile (2009a) and the present paper are distinguished from much of the prior literature by a neglect of estimation. This limited focus is shared by the concurrent work of Fox and Gandhi (2009), which explores identifiability of several related models, including a model of multinomial choice in which consumer types are multinomial and utility functions are analytic. More recently, Chiappori and Komunjer (2009) have explored identification of a multinomial choice model with an additive structure in a micro data environment, allowing for heterogeneous preferences and endogeneity through choice-specific unobservables. They use a change of variables approach more closely related to one of the approaches we have explored in Berry and Haile (2009a).

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" 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.

Each consumer i in market t is associated with a matrix of observables $z_{it} = (z_{i1t}, \ldots, z_{iJ_t t})$. The j subscript on z_{ijt} allows the possibility that some characteristics are both consumer- and choice-specific. This can result from interactions between consumer characteristics and product characteristics, or from measures that are inherently consumer/choice-specific. Examples from the literature include interactions between family size and automobile size (Goldberg (1995), Berry, Levinsohn, and Pakes (2004)), distance from an individual's home to a hospital, school, or retailer (Capps, Dranove, and Satterthwaite (2003), Hastings, Staiger, and Kane (2007), Burda, Harding, and Hausman (2008)), household exposure to product-specific advertising (Ackerberg (2001)), matches between a voter's position and party affiliation and those of each candidate (Rivers (1988)), and matches between the income, education and race of a household to that of a neighborhood (Bayer, Ferreira, and McMillan (2007)). We will require at least one such measure for each $j \geq 1$.[8]

In applications to demand it is important to model consumers as having the option to purchase none of the products the researcher focuses on.[9] 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 $J_t = |\mathcal{J}_t| - 1$.[10] Each inside good j has observable (to us) characteristics x_{jt}. Among other things, x_{jt} can include product dummies and price. Let $x_t = (x_{1t}, \ldots, x_{J_t t})$. Unobservables at the level of the product and/or market are characterized by a scalar $\xi_{jt}(z_{it})$ for each j, t, z_{it}. This may reflect unobserved choice characteristics, unobserved market characteristics, and/or unobserved taste for choice j in market t. Although we follow the literature in restricting $\xi_{jt}(z_{it})$ to be a scalar, we permit it to vary across subpopulations with different z_{it}. For simplicity we will assume that $\xi_{jt}(z_{it})$ has an atomless marginal distribution in the population for all z_{it}.

[8] If $z_{ijt} = \tilde{z}_{it}$ for all j, the identification problem is identical to that in the case of market-level data (see Berry and Haile (2009a)) conditional on each value of $\tilde{z}_{it}$.

[9] See, e.g., Bresnahan (1981), Anderson, DePalma, and Thisse (1992), Berry (1994), Berry, Levinsohn, and Pakes (1995), and Goldberg (1995).

[10] 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.

We consider preferences represented by a random utility model. Letting χ denote the support of $\left(x_{jt}, \xi_{jt}\left(z_{it}\right), z_{ijt}\right)$, 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. Given any $\left(x_{jt}, \xi_{jt}\left(z_{it}\right), z_{ijt}\right) \in \chi$,

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

where $\omega_{it} \in \Omega$, and u is measurable in ω_{it}. Without loss, the draw from the sample space Ω determining the realized function u_{it} is specified as independent of the arguments of the function, $\left(x_{jt}, \xi_{jt}\left(z_{it}\right), z_{ijt}\right)$; i.e., the measure $\mathbb{P}$ does not vary with $\left(x_{jt}, \xi_{jt}\left(z_{it}\right), z_{ijt}\right)$. This is without loss because the function u already permits the distribution of $u_{it}\left(x_{jt}, \xi_{jt}\left(z_{it}\right), z_{ijt}\right)$ to vary arbitrarily with $\left(x_{jt}, \xi_{jt}\left(z_{it}\right), z_{ijt}\right)$. Note that there is also no t subscript on the probability measure $\mathbb{P}$. This reflects our requirement that all unobserved choice/market-level heterogeneity be captured by $\xi_{jt}\left(z_{it}\right)$.

We let

$$v_{ijt} = u\left(x_{jt}, \xi_{jt}\left(z_{it}\right), z_{ijt}, \omega_{it}\right).$$

Despite the similarity of this notation to that in Matzkin (2007a) and Matzkin (2007b), we emphasize that here ω_{it} is not a random variable (or random vector) but an elementary event in the sample space Ω. The realization of ω_{it} could determine the realizations of an arbitrarily large number of random variables with arbitrary joint distribution. Using this observation, the following example shows how to map our general model to a more familiar special case.

Example 1. *Consider the linear random coefficients random utility model*

$$u\left(x_{jt},\xi_{jt},z_{ijt},\omega_{it}\right)=x_{jt}\beta_{it}+z_{ijt}\gamma+\xi_{jt}+\epsilon_{ijt}. \tag{3}$$

Here $\xi_{jt}\left(z_{it}\right)=\xi_{jt}$ *and 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* $\beta_{it}^{(k)}=\beta^{(k)}\left(\omega_{it}\right)$ *and* $\epsilon_{ijt}=\epsilon_j\left(\omega_{it}\right)$*.*[11] *This specification permits 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* t *and* $\left\{\left(x_{jt},\xi_{jt},z_{ijt}\right)\right\}_{j=1\ldots J}$*.*[12] *This specification of* $\left(\beta_{it},\left\{\epsilon_{ijt}\right\}_j\right)$ *is more general than typically allowed in the literature but more restrictive than required by our framework, even within a linear random coefficients model. For example, recalling that* x_t *can include product dummies, a more general model is obtained by letting* $\beta_{it}=\left(\beta^{(1)}\left(z_{it},\omega_{it}\right),\ldots,\beta^{(K)}\left(z_{it},\omega_{it}\right)\right)$ *and* $\epsilon_{ijt}=\epsilon\left(x_{jt},\xi_{jt},\omega_{it}\right)$*. We could generalize further by specifying* $\epsilon_{ijt}=\epsilon\left(x_{jt},\xi_{jt}\left(z_{it}\right),z_{ijt},\omega_{it}\right)$*; however, then the sum* $x_{jt}\beta_{it}+z_{ijt}\gamma+\xi_{jt}$ *in (3) would be redundant and the model would collapse to our completely general formulation (2).*

Aside from the restriction to scalar market/choice-specific unobservables, our representation of preferences is so far fully general. However, for most of our results we will restrict the specification by requiring separability in one component of z_{it}. Let $z_{ijt}=\left(z_{ijt}^{(1)},z_{ijt}^{(2)}\right)$, with $z_{ijt}^{(1)}\in\mathbb{R}$. Let $z_{it}^{(1)}$ denote the vector $\left(z_{i1t}^{(1)},\ldots,z_{iJ_tt}^{(1)}\right)'$ and $z_{it}^{(2)}$ the matrix $\left(z_{i1t}^{(2)},\ldots,z_{iJ_tt}^{(2)}\right)'$. We will require that for every $z_{it}^{(2)}$ there exist a representation of utilities with the form

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

where the function $\tilde{\mu}$ is strictly increasing in $\xi_{jt}\left(z_{it}^{(2)}\right)$, and the random coefficient $\phi_{it}=$

[11] The fact that x_t can include dummies for each product j permits us to write choice-specific functions such as $\epsilon_j\left(\cdot\right)$.

[12] 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 $\epsilon_{ijt}=\epsilon_j\left(\omega_{it}\right)$ for all possible choices j, not just those in the current choice set.

$\phi(\omega_{it})$ is strictly positive with with probability one.[13]

Here we have imposed three restrictions:

(i) *invariance* of $\xi_{jt}(z_{it})$ to $z_{it}^{(1)}$

(ii) *separability* in a "vertical" attribute $z_{ijt}^{(1)}$

(iii) *monotonicity* in ξ_{jt}.

Part (i) requires that there be one component of z_{it} whose variation is not confounded with variation in the unobservable. This is an important restriction although the standard assumption in the literature, that $\xi_{jt}(z_{it}) = \xi_{jt}\ \forall z_{it}$, is much stronger.[14] Part (ii) requires that, for each $z_{it}^{(2)}$, there be a monotonic transformation of utilities such that $z_{it}^{(1)}$ enters in additively separable form.[15] With separability, the specification of ω_{it} as independent of $x_{jt}, p_{jt}, \xi_{jt}(z_{it})$—previously without loss—now has bite: conditional on $\left(x_{jt}, \xi_{jt}, z_{it}^{(2)}\right)$, $z_{ijt}^{(1)}$ is independent of the stochastic component of $u_{it}\left(x_{jt}, \xi_{jt}, z_{ijt}\right)$. The invariance and separability restrictions together give $z_{ijt}^{(1)}$ the properties of a "special regressor" (e.g., Lewbel (2000)) that will provide a mapping between units of (latent) utility and units of (observable) choice probabilities.[16] Part (iii) requires monotonicity of utility in $\xi_{jt}\left(z_{it}^{(2)}\right)$. Because unobservables

---

[13] 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.

[14] Based on the results below, it would be easy to derive testable restrictions of this stronger assumption.

[15] For binary choice, if we assume $\xi_t(z_{it}) = \xi_t\left(z_{it}^{(2)}\right)$ and that $u(x_t, \xi_t(z_{it}), z_{it}, \omega_{it})$ is strictly increasing in $z_{it}^{(1)}$, then the event $\{u(x_t, \xi_t(z_{it}), z_{it}, \omega_{it}) > 0\}$ is equivalent to the event $\left\{z_{it}^{(1)} > u^{-1}\left(0; \xi_t\left(z_{it}^{(2)}\right), z_{it}^{(2)}, \omega_{it}\right)\right\}$. This leads to an observationally equivalent model with separability in $z_{it}^{(1)}$. This is well known. Nonetheless, additive separability is not without loss under these assumptions. This is because there may be no monotonic transformation of the original utility function that leads to the separable form. For example, suppose that according to $u\left(x_t, \xi_t\left(z_{it}^{(2)}\right), z_{it}, \omega_{it}\right)$ the marginal rate of substitution between $z_{it}^{(1)}$ and x_t varies with $z_{it}^{(1)}$. This property would be preserved by any monotonic transformation but fails under separability. An implication is that there can be simultaneous changes in $z_{it}^{(1)}$ and x_t that would raise welfare under one model but lower welfare under the other. Thus, although the separable structure preserves consumers' ordinal rankings of the outside good and any inside good, it need not preserve their ordinal rankings of alternative inside goods. Nonetheless, the observational equivalence demonstrates why it may be difficult to obtain full identification without a restriction like the separability we assume. Note that quasilinearity also provides a cardinal representation of utility, making it possible to characterize aggregate welfare.

[16] We can allow $z_{ijt}^{(1)}$ to be an index $g(c_{ijt})$ where c_{ijt} is a vector, as long as $\xi_{jt}\left(z_{it}^{(2)}\right) \perp\!\!\!\perp c_{ijt}$. Fixing t and $z_{it}^{(2)}$, utilities have the form $v_{ijt} = g(c_{ijt}) + \mu_{ijt}$ with $c_{ijt} \perp\!\!\!\perp \mu_{ijt}$. If $g(\cdot)$ is linear, identification of $g(\cdot)$ follows

have no natural order, monotonicity would be without loss 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 $\xi_{jt}\left(z_{it}^{(2)}\right)$ be a "vertical" rather than "horizontal" choice characteristic: all consumers agree that (all else equal) larger values of $\xi_{jt}\left(z_{it}^{(2)}\right)$ are preferred.[17] Of course, we allow heterogeneity in tastes for $\xi_{jt}\left(z_{it}^{(2)}\right)$ and for the vertical observable $z_{ijt}^{(1)}$. Furthermore, we allow a different representation (4) for each value of $z_{it}^{(2)}$. We show in section 7 that invariance, monotonicity in $z_{ijt}^{(1)}$, and monotonicity in $\xi_{jt}\left(z_{it}^{(2)}\right)$ all lead to testable implications.

Henceforth we will condition on $\mathcal{J}_t = \mathcal{J}$, with $|\mathcal{J}| = J$. We also condition on a value of $z_{it}^{(2)}$ and suppress this in the notation. For simplicity, we will now write only z_{ijt} and z_{it} to represent $z_{ijt}^{(1)}$ and $z_{it}^{(1)}$, respectively. Likewise, we will write ξ_{jt} instead of $\xi_{jt}\left(z_{it}^{(2)}\right)$.

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}, yielding utilities with the form

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

Here we include a subscript j on the function $\tilde{\mu}_j$ because in general we will have conditioned

by standard results (e.g., Manski (1985)). Identification of nonlinear $g(\cdot)$ can be obtained under restrictions considered in Matzkin (1993).

[17] Athey and Imbens (2007) point out that the assumption of a scalar vertical unobservable ξ_{jt} can lead to testable restrictions in some models. Although their observation does not apply to our more general model, below we derive a testable implication that does.

on a different value of $z_{ijt}^{(2)}$ for each j. Letting

$$\mu_j\left(x_{jt},\xi_{jt},\omega_{it}\right)=\frac{\tilde{\mu}_j\left(x_{jt},\xi_{jt},\omega_{it}\right)}{\phi_{it}}$$

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

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

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

3.3 Observables and Structural Features of Interest

Each consumer i maximizes her utility, choosing good j whenever $v_{ijt}>v_{ikt}\ \forall k\in\mathcal{J}-\{j\}$. For simplicity we assume that any ties with the outside good are broken in favor of the outside good. We denote consumer i's choice by

$$y_{it}=\arg\max_{j\in\mathcal{J}}\ v_{ijt}.$$

We will typically require excluded instruments, which we denote by $\tilde{\mathrm{w}}_{jt}$. One advantage of micro data is the variety of instrumental variables that may be available. Depending on the environment, instruments for prices[18] might include cost shifters excludable from the utility function, prices in other markets (e.g., Hausman (1996), Nevo (2001)), characteristics of competing products (e.g., Berry, Levinsohn, and Pakes (1995)), and/or "average" values of z_{it} in market t (e.g., Waldfogel (2003), Gentzkow and Shapiro (2007)).[19] Because the

[18] We discuss instruments for prices here because price is the leading case of an endogenous product characteristic. Our results permit any number/types of endogenous characteristics as long as valid instruments are available.

[19] The last of these obviously is unavailable without micro data. The analysis in Berry and Haile (2009a) suggests that characteristics of other goods may not be sufficient on their own without micro data or additional restrictions on preferences.

arguments are standard, we will not discuss assumptions necessary to justify the exclusion restrictions, which we will assume directly.

The observables then consist of

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

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

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

as known. Loosely speaking, we consider the case of observations from a large number of markets, each with a large number of consumers, who are anonymous conditional on z_{it}.

Our first objective is to derive sufficient conditions for identification of the market/choice-specific unobservables and the distribution of preferences over choices in sets $\mathcal{J}$, conditional on the characteristics $\left\{x_{jt}, z_{ijt}, \xi_{jt}\right\}_{j\in\mathcal{J}}$. In particular, we will show identification of $\left\{\xi_{jt}\right\}_{j\in\mathcal{J}}$ and of the joint distribution of $\{v_{ijt}\}_{j\in\mathcal{J}}$ conditional on any $\left\{x_{jt}, z_{ijt}, \xi_{jt}\right\}_{j\in\mathcal{J}}$ in their 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. Demand is fully characterized by the unobservables $\left\{\xi_{jt}\right\}_{j\in\mathcal{J}}$ and the

structural choice probabilities

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

These conditional probabilities are not directly observable from (6) because of the unobservables ξ_{jt}.

4 Illustration: Binary Choice with Exogenous Characteristics

To illustrate key elements of our approach, we begin with the simple case of binary choice with exogenous x_{jt}. Dropping the j subscripts, consumer i selects the inside good whenever

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

We consider identification under the following assumptions.

Assumption 1. $\xi_t \perp\!\!\!\perp (x_t, z_{it})$.

Assumption 2. supp $z_{it}|x_t = \mathbb{R}\ \forall x$.

Assumption 1 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 2 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)).[20] 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 and 2 are sufficient for full identification of the random

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

utility model. We can rewrite (5) as

$$v_{it} = z_{it} + \mu_{it} \tag{8}$$

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}. Since the observed conditional probability that a consumer chooses the outside good is

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

Assumption 2 then guarantees that $F_{\mu_{it}|t}(\cdot)$, the distribution of $\mu_{it}|t$, is identified from variation in z_{it} within market t. 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 and is a function only of x_t, and ξ_t; i.e.,

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

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 $\{\text{supp } x_t\} \times (0,1)$. 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

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

is uniquely determined for all $r \in \mathbb{R}$. Since $v_{it} = z_{it} + \mu_{it}$, this proves the following result.

Theorem 1. *Consider the binary choice setting with preferences given by (5). Under Assumptions 1 and 2, each ξ_t and the distribution of v_{it} conditional on any $(x_t, \xi_t, z_{it}) \in \chi$ is identified.*

Our argument involved two simple steps, each standard on its own. First, we showed that variation in z_{it} can be used to trace out the distribution of preferences across consumers. 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).[21] Here we apply this step *within each market*, exploiting the fact that conditioning on the market fixes all characteristics of the choice set, even those not observed. In the second step, we use variation in choice characteristics *across markets* to decompose the variation in utilities across products into the variation due to observables and that due to the unobservables ξ_t. This idea has been used extensively in estimation of parametric multinomial choice demand models following Berry (1994), Berry, Levinsohn, and Pakes (1995), and Berry, Levinsohn, and Pakes (2004). This second step is essential once we allow the possibility of endogenous choice characteristics (e.g., correlation between price and ξ_t), as will typically be necessary in demand estimation. Our approach for the more general cases follows the same broad outline.

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

5 Multinomial Choice: Full Identification

We now consider the general case of multinomial choice with endogenous characteristics using the specification of preferences in (5). We will use the following generalization of the large support assumption:

Assumption 3. $\text{supp}\left\{z_{ijt}\right\}_{j=1,\ldots,J} | \left\{x_{jt}\right\}_{j=1,\ldots,J} = \mathbb{R}^J$.

This is a strong assumption, essentially requiring sufficient variation in $(z_{i1t}, \ldots, z_{iJ_t t})$ to move choice probabilities through the entire unit simplex.[22] 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 1, we will require instrumental variables. 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, each taken from the recent literature on nonparametric instrumental variables regression.

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, with conditional density function $f_{x_j}\left(x_{jt}^{(1)} | \mathrm{w}_{jt}\right)$.[23] 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 $x_{jt}^{(1)}$. Let

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

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

[23] 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 focus on continuous endogenous characteristics here because price is our leading example.

and let $f_{\delta_j}(\cdot|x_{jt}, \mathrm{w}_{jt})$ denote its conditional density.

Let ϵ_τ and ϵ_f be some small positive constants. For each j and $\tau \in (0,1)$ define $\mathcal{L}_j(\tau)$ as the convex hull of functions $m_j(\cdot, \tau)$ that satisfy

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

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

We now assume the following instrumental variables conditions, from Chernozhukov and Hansen (2005, Appendix C).[24]

Assumption 4. $\xi_{jt} \perp\!\!\!\perp (\mathrm{w}_{jt}, z_{ijt}) \, \forall j, t$.

Assumption 5. For all j and $\tau \in (0,1)$, (i) 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 $\varepsilon_{jt} \equiv \delta_{jt} - D_j(x_{jt}, \tau)$, $E[B_j(x_{jt}, \tau)\psi_j(x_{jt}, \mathrm{w}_{jt}, \tau)|\mathrm{w}_{jt}] = 0$ a.s. only if $B_j(x_{jt}, \tau) = 0$ a.s., for $\psi_j(x, \mathrm{w}, \tau) = \int_0^1 f_{\varepsilon_j}(\sigma B_j(x, \tau)|x, \mathrm{w}) \, d\sigma > 0$.

(ii) the density $f_{\varepsilon_j}(e|x, \mathrm{w})$ of ϵ_{jt} is continuous and bounded in e over $\mathbb{R}$ a.s.;

(iii) $D_j(x_{jt}, \tau) \in s_j(x_{jt})$ for all x_{jt}.

Assumption 4 is a strong exclusion restriction requiring fully independent instruments. Assumption 5 is a particular type of "bounded completeness" condition, requiring 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 $\delta_j = D_j(x, \xi)$.[25] With these assumptions, we obtain the following result.

Theorem 2. *Under the representation of preferences in (5), suppose Assumptions 3–5 hold. Then each ξ_{jt} and the joint distribution of $\{v_{ijt}\}_{j \in \mathcal{J}}$ conditional on any $\{(x_{jt}, z_{ijt}, \xi_{jt})\}_{j \in \mathcal{J}}$ in their support is identified.*

[24] 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. Note that, as in their framework, δ_{jt} and x_{jt} can be assumed without loss to be tranformed such that they have bounded support.

[25] Chernozhukov and Hansen (2005) discuss sufficient conditions. We also consider an alternative to Assumption 5 below.

Proof. Let $\mu_{ijt} = \mu_j\left(x_{jt}, \xi_{jt}, \omega_{it}\right)$ and observe that[26]

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

Holding t fixed, $\mu_{ijt} \perp\!\!\!\perp z_{ijt}$ so that Assumption 3 guarantees identification of the marginal distribution of $\mu_{ijt}|t$ for each j. This implies identification of the conditional median

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

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

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

so that Assumption 3 implies identification of the joint distribution of $\left(\mu_{i1t}, \ldots, \mu_{iJt}\right)|t$. Since each x_{jt} is observed and each ξ_{jt} is identified, this implies identification of the joint distribution of $\left(\mu_{i1t}, \ldots, \mu_{iJt}\right)$ conditional on any $\left\{\left(x_{jt}, z_{ijt}, \xi_{jt}\right)\right\}_{j \in \mathcal{J}}$ in their support. Since $v_{ijt} = z_{ijt} + \mu_{ijt}$, the result follows. □

[26] Because we assume the large support condition from the beginning, our proof exploits this, using an "identification at infinity" argument commonly employed in this literature (e.g., Matzkin (1993)). We show in section 6 that this is not essential, however. See also Berry and Haile (2009a), which does not use an identification at infinity argument. They provide results for a market data setting, which is equivalent to a micro data setting with no variation in z_{it}.

5.2 Identification with Mean-Independent Instruments

Any application of instrumental variables methods requires that the instruments induce sufficient exogenous variation in the endogenous variables. Assumption 5 provides the appropriate formal condition for the model above, and this is the same condition that has been used to show identification of nonseparable regression models. Nonetheless, a possible limitation of Theorem 2 is that Assumption 5 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.[27] However, if we are willing to impose additional structure on the utility function, we can utilize a somewhat more intuitive sufficient condition for full identification. 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 as

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

where β_{it} and γ_{it} are strictly positive with probability one and the expectations $E\left[\beta_{it}\right]$, $E\left[\gamma_{it}\right]$, and $E\left[\tilde{\mu}_j \left(x_{jt}, \omega_{it}\right) | x_{jt}\right]$ are finite. This imposes a restriction relative to (5) but is still quite general relative to the prior literature. A representation of preferences equivalent to (13) is

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

where now

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

[27] In Berry and Haile (2009a) we explore an alternative argument relying on classical exclusion and support conditions in an environment combining discrete choice demand with a partially specified model of supply.

Here we will also 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.

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

Assumption 6. $E\left[\xi_{jt}|\mathrm{w}_{jt}, z_{ijt}\right] = 0 \ \forall j, t, \mathrm{w}_{jt}, z_{ijt}$.

To show identification of the joint distribution of $\{v_{ijt}\}_{j\in\mathcal{J}}$ conditional on $\left\{x_{jt}, z_{ijt}, \xi_{jt}\right\}_{j\in\mathcal{J}}$, first note that the marginal distribution of $\mu_{ijt}|t$ for each j is identified using the same argument used in the first lines of the proof of Theorem 2. This implies identification of the conditional means

$$\delta_{jt} \equiv E\left[\mu_j\left(x_{jt}, \xi_{jt}, \omega_{it}\right)\middle|\, t\right]$$

for all j and t. With the separable structure (15) and the normalization (16), for each j and t we have

$$\delta_{jt} = D_j\left(x_{jt}\right) + \xi_{jt}$$

for some unknown function D_j It is then straightforward to confirm that, under Assumption 6, 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 7. 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.

The completeness condition is the analog of the standard rank condition for identifiability in linear models. Like the standard rank condition, completeness requires that variation

in w_{ijt} induce sufficient variation in $x_{jt}^{(1)}$ to distinguish the true function $D_j(\cdot)$ from other functions of $x_{jt}^{(1)}$ using the conditional mean restriction of Assumption 6.[28]

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

Theorem 3. *Under the representation of preferences in (14)-(15), suppose Assumptions 3 and 6 hold. Then each ξ_{jt} and the joint distribution of $\{v_{ijt}\}_{j\in\mathcal{J}}$ conditional on any $\left\{(x_{jt}, z_{ijt}, \xi_{jt})\right\}_{j\in\mathcal{J}}$ in their support is identified if and only if Assumption 7 holds.*

Proof. From the preceding argument, under the completeness Assumption 7, 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). □

6 Identification of Demand Using Limited Support

The large support assumption (Assumption 3) in the preceding section is both common in the literature and controversial. Our results using this condition demonstrate that sufficient variation in the vector $(z_{i1}, \ldots, z_{iJ_t})$ 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 unobservables ξ_{jt} and the structural choice probabilities $\rho_j\left(\{x_{jt}, \xi_{jt}, z_{ijt}\}_{j\in\mathcal{J}}\right)$ at all points of support. We also show continuity of the identified features with respect to the support of

[28] Lehman and Romano (2005) give standard sufficient conditions and references. See also Newey and Powell (2003) and Severini and Tripathi (2006). If we add the assumption that $E\left[\tilde{\mu}_j\left(x_{jt}, \xi_{jt}\right)|x_{jt}\right]$ is bounded, the completeness assumption could be replaced with bounded completeness.

the micro data. In particular, moving from our limited support condition to the full support condition moves the identified features of the model smoothly toward the full identification results of the preceding sections.

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}, requiring a single "common choice probability" that is attainable in all markets by the appropriate choice of z_{it}.

Assumption 8. For some $q \in (0,1)$, for every market t there exists a unique $z_t^q \in$supp z_{it} such that $\Pr\left(y_{it}=1|z_{it}=z_t^q\right)=q$.

Here we require sufficient variation in z_{it} to push the choice probability to q in each market, not over the whole interval $(0,1)$ in each market.[29] 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). Here, the consumer chooses the inside good whenever (fixing $z_{it}^{(2)}$ and suppressing it)

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

[29]Implicitly we also require a continuous (region of) support for $\mu\left(x_t,\xi_t,\omega_i\right)|x_t,\xi_t$ to ensure uniqueness.

Under Assumption 8, for each market t we can find the z_t^q such that

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

Observe that each z_t^q is the qth 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

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

for some function $\zeta\left(\cdot; q\right)$ 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^q are those with low values of the unobservable ξ_t.

Identification of the function $\zeta\left(\cdot; q\right)$, 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 holds under the same type of bounded completeness assumption made in section 5.1; we state this condition formally as Assumption 12 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. This gives the following result.

Theorem 4. *In the binary choice model with preferences given by (5), suppose Assumptions 4, 8, and 12 hold. Then each ξ_t is identified and 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 $v_{it}|x_t, z_{it}, \xi_t$. In particular, we can identify a function $\zeta\left(\cdot; q\right)$ for each common choice probability q, each then determining

the qth quantile of $-\mu(x_t, \xi_t, \omega_{it})|x_t, \xi_t$. Since $v_{it} = z_{it} + \mu(x_t, \xi_t, \omega_{it})$, this determines the corresponding quantiles of the distribution of v_{it} conditional on (x_t, ξ_t, z_{it}). In the limit—i.e., with sufficient variation in z_{it} to make every $q \in (0,1)$ a common choice probability—all quantiles of v_{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 that the tails of z_{ijt} 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:[30]

$$v_{ijt} = \mu_j\left(z_{ijt} + \xi_{jt}, x_{jt}, \omega_{it}\right) \qquad \forall i, j = 1, \ldots, J \tag{19}$$

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

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

$$\lambda_{ijt} \equiv z_{ijt} + \xi_{jt}$$

rather than on each z_{ijt} and ξ_{jt} separately. Letting $\lambda_t = (\lambda_{i1t}, \ldots, \lambda_{iJt})$ and $x_t = (x_{1t}, \ldots, x_{Jt})$, we can then write the structural choice probabilities as

$$\rho_j(\lambda_t, x_t).$$

For the multinomial choice model our results rely on invertibility of the mapping from

[30] Here we focus on identification of demand. Continuity of the identified features can be obtained as in the binary model if we have the separable structure $z_{ijt} + \xi_{jt} + \mu_j(x_{jt}, \omega_{it})$.

the latent indexes λ_t to the market shares implied by utility maximization. To show that this holds we require a weak notion of the choice set $\mathcal{J}$ being a set of substitute goods that "belong" in the same market. To state this "connected substitutes" condition, we first need a definition.

Definition 1. *Product k substitutes to product ℓ at (λ_t, x_t)* if $\rho_\ell(\lambda_t, x_t)$ is strictly decreasing in λ_{ikt}.[31]

This definition provides a directional notion of one product being a substitute for another. For example, if a reduction in ξ_{kt} leads (all else equal) to a larger market share for product ℓ, we say that product k substitutes to product ℓ.

Given any values of (λ_t, x_t), let $\Sigma(\mathcal{J})$ denote the $(J+1) \times (J+1)$ matrix of zeros and ones, with the (r, c) element equal to one if product $(r-1)$ substitutes to product $(c-1)$ at (λ_t, x_t). We will assume that the products $j \in \mathcal{J}$ all belong in the same choice set in the following sense.

Assumption 9. At any (λ_t, x_t) such that $(\rho_1(\lambda_t, x_t), \ldots, \rho_J(\lambda_t, x_t))$ is on the interior of $\triangle^J$, the directed graph of $\Sigma(\mathcal{J})$ is strongly connected.

The directed graph of $\Sigma(\mathcal{J})$ has nodes (vertices) representing each product and an edge from product k to product ℓ whenever product k substitutes to product ℓ.[32] Our "connected substitutes" condition (Assumption 9) requires that there be a directed path from any product $j \in \mathcal{J}$ to any other product $j' \in \mathcal{J}$. Of course, the path between j and j' may be through other nodes. Thus, for example, even a market with two independent goods satisfies this condition if both substitute to and from the outside good. On the other hand, if the connected substitutes condition fails, then there is some strict subset of $\mathcal{J}$ that substitute only among themselves for at least some values of (λ_t, x_t). In Berry and Haile

[31] Because we introduce this assumption after normalizing the utility of the outside good to zero, we define an increase in λ_{i0t} to mean equal reductions in λ_{ijt} for all $j > 0$. Thus product 0 substitutes to product j if the probability good j is selected goes up whenever λ_{ikt} increases by an equal amount for all $k > 0$.

[32] In standard models $\Sigma(\mathcal{J})$ will typically be symmetric, so the edges of its directed graph will be bi-directional.

(2009a) we provide additional discussion of this condition and show that it is satisfied in standard models. There we also point out that this condition is equivalent to a condition used by Gandhi (2008) to show invertibility of market shares. Using his argument, we can show the following result for our framework, generalizing well-known invertibility results for linear discrete choice models in Hotz and Miller (1993), Berry (1994) and Berry and Pakes (2007).[33]

Lemma 1. *Consider any choice probability vector* $p = (p_1, \ldots, p_J)'$ *on the interior of* $\triangle^J$. *Under Assumption 9, for any* x_t *there is at most one vector* $\lambda \in \mathbb{R}^J$ *such that* $\rho_j(\lambda, x_t) = p_j$ *for all* j.

Proof. See Berry and Haile (2009a). □

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

Assumption 10. There exists $q = (q_0, q_1, \ldots, q_J)$ on the interior of $\triangle^J$ such that for every market t there is a vector $z_t^q = (z_{1t}^q, \ldots, z_{1t}^q) \in \text{supp}(z_{i1t}, \ldots, z_{iJ_t t})$ such that for all j, $q_j = \Pr(y_{it} = j \mid x_{1t}, \ldots, x_{Jt}, z_{i1t}, \ldots, z_{iJ_t t})_{z_{it}=z_t^q}$.

Assumption 10 requires the vector $(z_{i1t}, \ldots, z_{iJt})$ 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 still requires sufficiently rich J-dimensional micro data; however, it is considerably weaker than the full support condition, which essentially required all elements of $\triangle^J$ to be common choice probabilities.

[33] Berry (1994) and Berry and Pakes (2007) show existence and uniqueness of an inverse choice probability in models with an additive structure. Gandhi (2008) relaxes the separability requirement. Our lemma addresses only uniqueness conditional on existence. Under the maintained assumption that the model is correctly specified, any observed choice probability vector must have a vector $(\lambda_1, \ldots, \lambda_J)$ that rationalizes it. Gandhi (2008) provides additional conditions guaranteeing 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 somewhat more general model of preferences.

Theorem 5. *In the multinomial choice model with preferences given by (19), suppose Assumptions 6, 7, 9, and 10 hold. Then each ξ_{jt} is identified and the structural choice probabilities $\rho_j\left(\{x_{jt}, \xi_{jt}, z_{ijt}\}_{j\in\mathcal{J}}\right)$ are identified at all $\{x_{jt}, \xi_{jt}, z_{ijt}\}_{j\in\mathcal{J}}$ in their support.*

Proof. Fix $x_t = (x_{1t}, \ldots, x_{Jt})$ and let q be the common choice probability vector. By Lemma 1, there is a unique vector $\lambda(x_t, q) = (\lambda_1(x_t, q)., \ldots, \lambda_J(x_t, q))$ such that

$$\rho_j(\lambda(x_t, q), x_t) = q_j \quad \forall j.$$

Further, by the definition of z_t^q and $\lambda_j(x_t, q)$, $\lambda_j(x_t, q) = \xi_{jt} + z_{jt}^q$, so that

$$z_{jt}^q = \lambda_j(x_t, q) - \xi_{jt} \quad \forall j, t. \tag{20}$$

Under Assumptions 6 and 7, the equations (20) identify the functions $\lambda_j(\cdot, q)$ and each ξ_{jt} for all j and t, 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. □

Note that, in contrast to the results in section 5, here the entire x_t vector appears as an argument of the "regression" function λ_j in (20). One implication is that characteristics of competing products are not available as instruments.

7 Testable Restrictions

The models we have considered incorporate two important maintained assumptions: (i) existence of a vertical consumer-choice observable z_{ijt}; (ii) a scalar vertical market/choice-specific unobservable, ξ_{jt}. Here we show that these assumptions imply testable restrictions.

The following remark points out that the assumption of a vertical z_{ijt} has immediate implications for the observed conditional choice probabilities: choice probabilities must increase in z_{ijt}.

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

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

The assumption of a scalar vertical unobservable also leads to testable implications. For simplicity we show this here for binary choice. To state the result it will be useful to recall Theorem 4 and let $\xi_t(z_t^q; q, x_t)$ denote the value of ξ_t identified from the common choice probability q in market t.

Remark 2. *In the binary choice model with preferences given by (5), suppose Assumptions 4, 12, and 8 hold. Then $\xi_t(z_t^q; q, x_t)$ must be strictly decreasing in z_t^q across markets.*

This follows from the fact that v_{it} is strictly increasing in both z_{it} and ξ_t under the assumptions of the model. Thus, the value z_t^q required to attain the common choice probability q in market t will be higher when the unobservable ξ_t is lower. 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 2. *Suppose $\mu(x_t, \xi_t, \phi_{it}) = -\nu_{it}\xi_t$, with $\nu_{it} \sim N(0,1)$. Take $q > 1/2$ and consider the set of markets in which $\xi_t(z_t^q; q, x_t) > 0$. Recall that each z_t^q is observable and is defined such that $\Pr(\nu_{it}\xi_t < z_t^q) = q$. Letting Φ denote the standard normal CDF, this requires*

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

Therefore, by construction, $\frac{z_t^q}{\xi_t}$ *will take the same value in every market. Since each* z_t^q *must also be positive when* $q > 1/2$*, this requires a strictly positive correspondence between* z_t^q *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 ξ_{jt} that is invariant to z_{it} : the values of ξ_{jt} inferred from any two common choice probabilities must agree.

Remark 3. *In the binary choice model with preferences given by (5), suppose Assumptions 4 and 12 hold. In addition, suppose that for distinct* q *and* q' *in the interval* $(0,1)$*, for every market* t *there exists a unique* $z_t^q \in$*supp* z_{it} *such that* $\Pr\left(y_{it}=1|z_{it}=z_t^q\right)=q$ *and a unique* $z_t^{q'} \in$*supp* z_{it} *such that* $\Pr\left(y_{it}=1|z_{it}=z_t^{q'}\right)=q'$*. Then* $\xi_t\left(z_t^q;q,x_t\right)=\xi_t\left(z_t^{q'};q',x_t\right)$ *for all* t*.*

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

The hypothesis of Remark 3 merely requires that there be two distinct common quantiles. The following example demonstrates that the resulting restriction can fail if the assumption of a scalar unobservable is violated.

Example 3. *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* $z_{it}=z_t^q$ *only consumers with* $\nu_{it}>1-q$ *choose the inside good. Thus, the value of* z_t^q *is determined by the preferences of the consumer with* $\nu_{it}=1-q$*. Now consider the* $\xi_t(q)$ *inferred under the incorrect assumption of a scalar unobservable. From the observations above, when* $q>1/2$ *we have* $\xi_t(q)=F_{\xi^1+\xi^2}\left(\xi_t^1+\xi_t^2\right)$ *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, $\xi_t(q)$ will equal σ. Similarly, for $q' < 1/2$, $\xi_t(q') = F^{-1}_{\xi^1+2\xi^2}\left(\xi_t^1 + 2\xi_t^2\right)$; i.e., if $\xi_t^1 + 2\xi_t^2$ fall at the σ' quantile of this sum in the cross section of markets, $\xi_t(q')$ will be σ'. In general, $\sigma \neq \sigma'$.

8 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 (2009a) 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

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

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 11. supp $(x_{1t}^{(i)}, \ldots, x_{Jt}^{(i)}) | (x_{1t}^{(ii)}, \ldots, x_{Jt}^{(ii)}) = \mathbb{R}^J \ \forall t$.

Assumption 11 is different from the parallel large support Assumption 3. 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). Here, however, the notion of "market group" replaces that of "market." Within a market group, $x_{jt}^{(i)}$ now plays the role of the micro data.

With this reinterpretation, the setup is isomorphic to that in section 5, and the prior arguments apply directly. Variation in $x_{jt}^{(i)}$ across market groups at the limit $x_{j't}^{(i)} \to -\infty \, \forall j' \neq j$ identifies the distribution of $\mu_i\left(x_{j\Gamma}^{(ii)}, \xi_{j\Gamma}\right)$ exactly as in section 5. Letting $\delta\left(x_{j\Gamma}^{(ii)}, \xi_{j\Gamma}\right) = E\left[\mu_i\left(x_{j\Gamma}^{(ii)}, \xi_{j\Gamma}\right) | x_{j\Gamma}^{(ii)}, \xi_{j\Gamma}\right]$, identification of the function $\delta\left(x_{j\Gamma}^{(ii)}, \xi_{j\Gamma}\right)$ (and therefore each $\xi_{j\Gamma}$) follows exactly as in the previous sections. With each $\xi_{j\Gamma}$ and the distribution of $\mu_i\left(x_{j\Gamma}^{(ii)}, \xi_{j\Gamma}\right)$ known, the conditional joint distribution of $\{v_{ijt}\}_{j\in\mathcal{J}} \left| \left(\mathcal{J}, \left\{(x_{jt}, \xi_{jt})\right\}_{j\in\mathcal{J}}\right)\right.$ is uniquely determined at any $\left(\mathcal{J}, \left\{(x_{jt}, \xi_{jt})\right\}_{j\in\mathcal{J}}\right)$ in their support.

Because the setup here is isomorphic to that for the original micro data setting, extensions to the case of endogenous characteristics (elements of $x_{jt}^{(ii)}$), a separable error structure, and identification of demand with limited support follow directly as well.[34]

9 Conclusion

We have studied nonparametric identification of models of multinomial choice demand, allowing for market/choice-specific unobservables, endogenous choice characteristics, and arbitrary random heterogeneity across consumers in tastes for products and/or characteristics.

[34] An interesting but unresolved 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

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

For any finite J, 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. As $J \to \infty$ one would then obtain an arbitrarily good approximation of the joint distribution of (δ_j, x_j). If this joint distribution were instead *known*, the fact that $\delta_j = D\left(x_j, \xi_j\right)$ would allow identification of D from the results of Chernozhukov and Hansen (2005). Identification of demand and full identification would then follow.

We obtained full identification using the same kind of large support assumption used to show identification in even the simplest semiparametric discrete choice 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, whereas 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. 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 observables.

A second limitation concerns welfare. Our model (5) incorporates quasilinear preferences. This provides a specification of cardinal utility that can be used to characterize changes in utilitarian social welfare (in aggregate, or across subpopulations defined by observables)[35] or changes in welfare under any social welfare function that is anonymous conditional on observables. However, our model lacks the structure required for welfare analysis that depends on the distribution of changes in individual utilities. Characterization of Pareto improvements, for example, would require tracking each individual consumer's position in the distribution of utilities before a policy change to that after. Our model specifies a distribution of conditional indirect utilities, not a distribution of parameters whose realizations can be associated with individual consumers. This points out a limitation of nonparametric random utility models as a theoretical foundation for some kinds of welfare analysis: such

[35]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 individuals' positions in the distributions of normalized utilities as the choice environment varies, prevents characterization of aggregate compensating variation or equivalent variation in income units.

welfare calculations require additional *a priori* structure.

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

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

This generates a special case of our model, so we have provided conditions for identification of $\{\xi_{jt}\}_{j\in\mathcal{J}}$ and the joint distribution of $\{v_{ijt}\}_{j\in\mathcal{J}} \,|\, \{x_{jt}, \xi_{jt}, z_{ijt}\}_{j\in\mathcal{J}}$. However, it is clear that the joint distribution of $(\beta_{it}, \epsilon_{i1t}, \ldots, \epsilon_{iJt})$ is not identified without additional restrictions.[36] Going from our results to identification of the distribution of parameters in (23) is equivalent to the standard problem of 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 pursuing this line of argument 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)) is an open question.

Finally, while a novel aspect of our work is its 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, and we explore that case in Berry and Haile (2009a). Whether the sufficient conditions for identification 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.

[36] For any true model with the form (23), an observationally equivalent model is obtained by setting $\beta_{it} = \gamma = 0$ and letting the joint distribution of $(\epsilon_{i1t}, \ldots, \epsilon_{iJt}) \,|\, \{x_{jt}, z_{ijt}, \xi_{jt}\}_{j\in\mathcal{J}}$ equal that of $(v_{i1t} - \xi_{1t}, \ldots, v_{iJt} - \xi_{Jt}) \,|\, \{x_{jt}, z_{ijt}, \xi_{jt}\}_{j\in\mathcal{J}}$.

Appendix

Here we state Assumption 12, used in Theorem 4. From equation (17) we have

$$z_t^q = \zeta\left(x_t, \xi_t; q\right)$$

where x_t denotes the endogenous characteristic of choice 1. Let $f_{z^q}\left(\cdot|x_t, \mathrm{w}_t\right)$ denote the density of z_t^q conditional on x_t and the instruments w_t. Fix some small positive constants $\epsilon_\tau, \epsilon_f > 0$. Fix $q \in (0,1)$. 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, $\Pr\left(z_t^q \leq m\left(x_t, \tau\right)|\mathrm{w}_t\right) \in [\tau - \epsilon_\tau, \tau + \epsilon_\tau]$; and

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

Assumption 12. For all $\tau \in (0,1)$, (i) for any bounded function $B(x, \tau) = m(x, \tau) - \zeta(x, \tau; q)$ with $m(\cdot, \tau) \in \mathcal{L}(\tau)$ and $\varepsilon_t \equiv z_t^q - \zeta\left(x_t, \tau; q\right)$, $E\left[B\left(x_t, \tau\right)\psi\left(x_t, \mathrm{w}_t, \tau\right)|\mathrm{w}_t\right] = 0$ a.s. only if $B\left(x_t, \tau\right) = 0$ a.s., for $\psi\left(x, \mathrm{w}, \tau\right) = \int_0^1 f_\varepsilon\left(\sigma B\left(x, \tau\right)|x, \mathrm{w}\right)\, d\sigma$.

(iii) the density $f_\varepsilon\left(e|x, \mathrm{w}\right)$ of ϵ_t is continuous and bounded in e over $\mathbb{R}$ a.s.;

(iv) $\zeta\left(x_t, \tau; q\right) \in s\left(x_t\right)$ for all x_t.

References

Ackerberg, D. A. (2001): "Empirically Distinguishing Informative and Prestige Effects of Advertising," *RAND Journal of Economics*, 32(2), 316–333.

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

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

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

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

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

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

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

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

Berry, S. T., and P. A. Haile (2009a): "Identification in Differentiated Products Markets Using Market Level Data," Discussion paper, Yale University.

——— (2009b): "Identification of a Nonparametrc Generalized Regression Model with Group Effects," Discussion paper, Yale University.

Berry, S. T., and A. Pakes (2007): "The Pure Characteristics Demand Model," Discussion paper, Yale University.

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.

Burda, M., M. Harding, and J. Hausman (2008): "A Bayesian Mixed Logit-Probit Model for Multinomial Choice," Discussion paper, M.I.T.

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.

Chiappori, P.-A., and I. Komunjer (2009): "On the Nonparametric Identification of Multiple Choice Models," Discussion paper, University of California San Diego.

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

Fox, J., and A. Gandhi (2009): "Identifying Heterogeneity in Economic Choice Models," 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.

Gentzkow, M., and J. Shapiro (2007): "What Drives Media Slant," Discussion paper, University of Chicago.

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., and D. Wise (1978): "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences," *Econometrica*, 46, 403–426.

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.

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.

Hotz, J., and R. A. Miller (1993): "Conditional Choice Probabilites and the Estimation of Dynamic Models," *Review of Economic Studies*, 60, 497–529.

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.

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.

MANSKI, C. F., AND D. A. WISE (1983): *College choice in America*. Harvard University Press, Cambridge, MA.

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.

Poole, K. T., and H. Rosenthal (1985): "A Spatial Model for Legislative Roll Call Analysis," *American Journal of Political Science*, 29(2), 357–384.

Quandt, R. E. (1966): "A Probabalistic Abstract Mode Model," in *Studies in Travel Demand, Volume II*, pp. 90–113. Mathematica, Princeton, N.J.

——— (1968): "Estimation of Modal Splits," *Transportation Research*, 2, 41–50.

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

Schultz, T. P. (1982): "Lifetime Migration within Educational Strata in Venezuela: Estimates of a Logistic Model," *Economic Development and Cultural Change*, 30(3), 559–593.

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

Waldfogel, J. (2003): "Preference Externalities: An Empirical Study of Who Benefits Whom in Differentiated-Product Markets," *RAND Journal of Economics*, 34(3), 557–568.